

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

ПОЯСНЮВАЛЬНА ЗАПИСКА

до кваліфікаційної роботи освітнього ступеня «магістр»
за спеціальністю 121 «Інженерія програмного забезпечення»
(освітня програма «Інженерія програмного забезпечення»)

на тему:

**«Аналіз даних щодо потреб та відвідуваності
клієнтів адвокатського бюро»**

Виконав студент групи ПЗм-22-1

МЕЛЬНИК Роман Ігорович

Керівник роботи:

КРАВЧЕНКО Світлана Миколаївна

Рецензент:

ЛОКТИКОВА Тамара Миколаївна

Житомир – 2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

«ЗАТВЕРДЖУЮ»

Зав. кафедри інженерії
програмного забезпечення

Тетяна ВАКАЛЮК

«18» вересня 2023 р.

ЗАВДАННЯ
на кваліфікаційну роботу

Здобувач вищої освіти: **МЕЛЬНИК Роман Ігорович**

Керівник роботи: **КРАВЧЕНКО Світлана Миколаївна**

Тема роботи: **«Аналіз даних щодо потреб та відвідуваності клієнтів
адвокатського бюро»,**

затверджена Наказом закладу вищої освіти від **«18» вересня 2023 р. № 479/с**

Вихідні дані для роботи: **предметом дослідження є моделі й методи аналізу
потреб та відвідуваності користувачів. Об'єктом дослідження є
застосування сучасних технологій та аналіз їх властивостей.**

Консультанти з магістерської кваліфікаційної роботи із зазначенням розділів,
що їх стосуються:

Розділ	Консультант	Завдання видав	Завдання прийняв
1	<i>Кравченко С.М.</i>	22.09.2023р.	22.09.2023р.
2	<i>Кравченко С.М.</i>	22.09.2023р.	22.09.2023р.
3	<i>Кравченко С.М.</i>	22.09.2023р.	22.09.2023р.

Дата видачі завдання: **«22» вересня 2023 р.**

РЕФЕРАТ

Кваліфікаційна магістерська робота являє собою наукове дослідження методів та моделей аналізу даних про потреби й поведінку користувачів веб-ресурсу на прикладі сайту адвокатського бюро.

Робота на тему «Аналіз даних щодо потреб та відвідуваності клієнтів адвокатського бюро» включає теоретичне обґрунтування та практичну реалізацію комплексу засобів для дослідження активності користувачів сайту, а також пояснювальну записку.

Пояснювальна записка до магістерської роботи містить 64 сторінки, 17 ілюстрацій та 1 таблицю.

У роботі сформульовано основні задачі з аналізу веб-аналітики, проаналізовано існуючі підходи та інструменти, обґрунтовано вибір оптимальних технологій і методів дослідження на основі задач конкретного проєкту.

КЛЮЧОВІ СЛОВА: ВЕБ-АНАЛІТИКА, DATA SCIENCE, ІНТЕРНЕТ-МАРКЕТИНГ, ВІДВІДУВАНІСТЬ САЙТУ, АНАЛІЗ ДАНИХ, МОДЕЛЮВАННЯ, ОПТИМІЗАЦІЯ САЙТУ.

ABSTARCT

The qualifying master's thesis is a scientific research of methods and models for analyzing data on the needs and behavior of website users on the example of a law firm website.

The thesis on "Analysis of data on the needs and attendance of law firm clients" includes a theoretical justification and practical implementation of a set of tools for studying user activity on the website, as well as an explanatory note.

The explanatory note to the master's thesis contains 64 pages, 17 illustrations and 1 table.

The paper formulates the main tasks for web analytics analysis, analyzes existing approaches and tools, and substantiates the choice of optimal technologies and research methods based on the objectives of a particular project.

KEYWORDS: WEB ANALYTICS, DATA SCIENCE, INTERNET MARKETING, WEBSITE VISITORS, DATA ANALYSIS, MODELING, WEBSITE OPTIMIZATION.

ЗМІСТ

РЕФЕРАТ	3
ВСТУП.....	7
РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМАТИКИ, МЕТОДІВ ТА ЗАСОБІВ ВИРІШЕННЯ ЗАДАЧІ	10
1.1. <i>Постановка завдання дослідження</i>	10
1.2. <i>Оцінка потенційних труднощів та викликів</i>	16
1.3. <i>Порівняльний огляд існуючих досліджень і програмних рішень.</i>	18
1.4. <i>Обґрунтування вибору інструментальних засобів та визначення вимог до реалізації програмного продукту.....</i>	25
1.5. <i>Ключові аспекти методів аналізу потреб та відвідуваності</i>	27
<i>Висновки до першого розділу</i>	33
РОЗДІЛ 2. РОЗРОБКА МЕТОДІВ АНАЛІЗУ ДАНИХ ПРО ВІДВІДУВАЧІВ САЙТУ	35
2.1 <i>Збір та підготовка даних</i>	35
2.2 <i>Аналіз якісних характеристик відвідувачів.....</i>	39
2.3 <i>Побудова прогностичних моделей відвідуваності.....</i>	43
2.4 <i>Реалізація алгоритмів роботи системи.....</i>	50
<i>Висновки до другого розділу</i>	57
РОЗДІЛ 3. РОЗРОБКА РЕКОМЕНДАЦІЙ ДЛЯ ОПТИМІЗАЦІЇ САЙТУ	59
3.1 <i>Виявлення UX проблем та потреб користувачів.....</i>	59
3.2 <i>Формування рекомендацій для оптимізації</i>	60
3.3 <i>Оцінка ефективності запропонованих рішень.....</i>	61
<i>Висновки до третього розділу.....</i>	62
ВИСНОВКИ	64
СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ	65
ДОДАТКИ	69

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

АД	— аналіз даних.
МН	— машинне навчання
ВР	— відвідуваність ресурсу
ВА	— веб-аналітика
ІМ	— інтернет-маркетинг
КЦ	— коефіцієнт конверсії
UX	— user experience (користувацький досвід)
UI	— user interface (користувацький інтерфейс)
РСК	— рекомендації та стратегії клієнтів
GA	— Google Analytics

ВСТУП

Актуальність теми. Аналіз даних щодо потреб та відвідуваності клієнтів сайту адвокатського бюро є вкрай актуальним в умовах стрімкого розвитку цифрових технологій та їх проникнення в усі сфери економічної діяльності, включаючи юридичний бізнес.

Ефективна присутність в мережі Інтернет, оптимізація каналів онлайн-маркетингу, підвищення відвідуваності сайту та рівня конверсій, створення зручного користувацького досвіду взаємодії з сервісами компанії стає ключовим фактором конкурентоспроможності сучасних адвокатських фірм.

З огляду на це, дослідження поведінки користувачів на сайтах юридичних компаній, вивчення їх потреб, уподобань, можливих труднощів під час отримання послуг, оцінка ефективності бізнес-процесів з точки зору клієнтського досвіду є вкрай актуальними науково-прикладними задачами.

Їх вирішення дозволить підвищити якість обслуговування клієнтів, прискорити їх обробку та оптимізувати взаємодію з адвокатською фірмою в мережі Інтернет. Це, в свою чергу, сприятиме збільшенню потоку нових клієнтів, формуванню їх loyalty та зміцненню конкурентних переваг компанії.

Для вирішення подібних прикладних задач потрібні нові наукові підходи на основі сучасного інструментарію аналізу даних, машинного навчання, прогнозної аналітики, які дозволять знаходити неочевидні закономірності у взаємодії із споживачами та приймати обґрунтовані рішення щодо оптимізації бізнесу.

Саме тому розробка методів збору та аналізу даних про поведінку користувачів юридичних сайтів, побудова моделей передбачення їх потреб та відвідуваності ресурсу, формулювання науково обґрунтованих рекомендацій для оптимізації UX/UI сайту та маркетингових бізнес-процесів є вкрай затребуваною та актуальною задачею для сфери юридичних послуг.

Мета магістерської роботи полягає в ретельному та системному аналізі даних, пов'язаних із потребами та відвідуваністю клієнтів в адвокатському бюро. Основний акцент роботи спрямований на використання методів обробки даних

та статистичного аналізу для вивчення змін у виборі юридичних послуг та виявлення ключових факторів, які впливають на клієнтські рішення.

Результати роботи дозволять розробити рекомендації для оптимізації стратегії обслуговування клієнтів, вдосконалення послуг та підвищення конкурентоспроможності адвокатського бюро на ринку юридичних послуг.

Завдання. Для досягнення цієї мети робота включає наступні завдання:

1. Проаналізувати існуючі підходи та літературні джерела з тематики дослідження. Обґрунтувати вибір методів аналізу даних.

2. Зібрати необхідні дані про відвідувачів сайту. Здійснити їх попередню обробку та аналіз за допомогою обраних методів. Побудувати прогнознi моделі динаміки відвідуваності.

3. На основі отриманих результатів аналізу розробити конкретні рекомендації для вдосконалення роботи сайту. Запропонувати способи оцінки ефективності реалізованих рішень.

4. Реалізувати програмний код для автоматизації процесів збору та аналізу даних про відвідувачів сайту.

5. Оформити результати дослідження у вигляді магістерської роботи відповідно до встановлених вимог.

Об'єктом дослідження є моделі та методи для аналізу даних потреб та відвідуваності клієнтів адвокатського бюро.

Предмет дослідження: побудова аналітичних моделей, ідентифікація ключових факторів взаємодії та вивчення можливостей оптимізації через дані та аналітику.

Практичне значення:

- Розроблені методи аналізу відвідуваності веб-ресурсу дозволять оптимізувати маркетинг та бізнес-процеси компанії на основі даних про поведінку цільової аудиторії.

- Запропоновані рекомендації щодо оптимізації юзабіліті та UX дизайну сайту сприятимуть покращенню користувацького досвіду та збільшенню конверсій сайту.

- Розроблений інструментарій збору, аналізу та моделювання даних про відвідувачів може застосовуватися для оптимізації будь-яких інших сайтів та веб-ресурсів.

- Створені аналітичні моделі та панелі візуалізації дозволять зацікавленим особам приймати обґрунтовані рішення на основі даних.

- Отримані результати можуть слугувати основою для подальших наукових досліджень в області веб-аналітики та інтернет-маркетингу.

Апробація результатів дослідження. Матеріали кваліфікаційної роботи доповідалися на Тези VI Всеукраїнської науково-технічної конференції «Комп'ютерні технології: інновації, проблеми, рішення» [41, 42].

РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМАТИКИ, МЕТОДІВ ТА ЗАСОБІВ ВИРІШЕННЯ ЗАДАЧІ

1.1. Постановка завдання дослідження

Аналіз даних щодо потреб та відвідуваності адвокатського бюро — це процес вивчення взаємодії клієнтів із юридичною фірмою з метою забезпечення кращого розуміння їхніх потреб і покращення якості юридичних послуг. Такий аналіз включає в себе дослідження різних аспектів, таких як частота відвідування клієнтів, їхні основні потреби, задоволеність від співпраці та можливість надання додаткових юридичних послуг.

Аналіз даних про відвідувачів веб-сайтів є надзвичайно актуальною задачею в сучасних умовах. Адвокатські контори, зокрема, зацікавлені у вивченні поведінки користувачів на своїх сайтах, аби краще зрозуміти їхні потреби та очікування і, відповідно, покращити якість обслуговування.

Отже, дослідження даних щодо відвідуваності та активності користувачів на сайті конкретної адвокатської фірми є вкрай необхідним для вдосконалення її роботи в мережі Інтернет. Аналізуючи дані про перегляди сторінок, завантаження файлів, заповнення форм та інших дій користувачів, можна зрозуміти слабкі та сильні сторони ресурсу, його недоліки в юзабіліті, а також дізнатися більше про цільову аудиторію сайту.

Мета дослідження — проаналізувати дані щодо відвідуваності сайту конкретної адвокатської фірми за певний період, виявити закономірності у поведінці користувачів, знайти проблемні місця ресурсу та надати рекомендації щодо його вдосконалення.

В роботі будуть використані такі методи як збір даних, їх первинна обробка, аналіз, візуалізація, моделювання. Серед інструментів слід виділити Google Analytics, Microsoft Excel, Tableau, Python (бібліотеки Pandas, Matplotlib, Seaborn).

Етапи дослідження:

1. Збір даних з Google Analytics

2. Первинна обробка та очищення даних в Excel
3. Завантаження очищених даних в Python
4. Аналіз та візуалізація даних за допомогою Pandas, Matplotlib, Seaborn
5. Побудова прогнозних моделей в Python
6. Інтерпретація результатів, формулювання висновків
7. Підготовка презентації результатів у Tableau
8. Надання рекомендацій щодо вдосконалення сайту

В підсумку ми очікуємо отримати детальний аналіз відвідуваності сайту адвокатської фірми, виявити сильні та слабкі сторони ресурсу, зрозуміти уподобання цільової аудиторії. Також будуть побудовані прогнозні моделі подальшої поведінки користувачів та надані конкретні рекомендації щодо оптимізації сайту. Це дозволить покращити конверсію ресурсу та якість обслуговування клієнтів адвокатської фірми.

Для початку, важливо розглянути роль соціальних відносин у сфері юридичних послуг. Аналіз взаємодії клієнтів з адвокатським бюро в контексті соціальних мереж може включати в себе оцінку активності клієнтів у віртуальних спільнотах, обговорення юридичних питань у Інтернеті, а також їхню участь у відгуках та рекомендаціях.

Адвокатське бюро може також створювати індивідуальні профілі для клієнтів, щоб отримати більше інформації про їхні потреби та попередні юридичні консультації. Це дозволить більш ефективно аналізувати дані та персоналізувати надані послуги.

Основною метою аналізу є виявлення ключових аспектів, які впливають на задоволеність клієнтів та їхню відданість адвокатському бюро. Пошук зв'язків між відвідуваністю, видами послуг та задоволеністю клієнтів є критичним для оптимізації роботи юридичної фірми та розвитку стратегій вдосконалення обслуговування.

Наприклад, можливий аспект аналізу - це оцінка популярності різних юридичних послуг серед клієнтів та їхнє сприйняття якості цих послуг. За

результатами аналізу можна розробити рекомендації щодо покращення надання певних видів юридичних консультацій або розширення спектру послуг.

Також, важливо розглядати аспекти конфіденційності та безпеки даних клієнтів. Забезпечення високого рівня захисту інформації та дотримання всіх норм та законодавства є невід'ємною частиною такого дослідження. Аналіз даних щодо потреб та відвідуваності адвокатського бюро дозволяє створити ефективні стратегії для залучення та утримання клієнтів, а також постійно вдосконалювати якість наданих юридичних послуг. Аналіз даних щодо потреб та відвідуваності адвокатського бюро розглядається через призму різних аспектів, що визначають взаємодію клієнтів із юридичною фірмою. Даний аналіз можна розділити на кілька типів залежно від специфіки дослідження:

1. Частота відвідування клієнтами адвокатського бюро:
 - Оцінка регулярності відвідування клієнтами адвокатської фірми.
 - Визначення факторів, що впливають на збереження постійних клієнтів та залучення нових.
2. Види послуг, якими користуються клієнти:
 - Дослідження популярності різних юридичних послуг серед клієнтів.
 - Аналіз вибору клієнтами конкретних послуг та їхнього сприйняття якості.
3. Оцінка задоволеності клієнтів:
 - Вивчення відгуків та рейтингів клієнтів стосовно наданих послуг.
 - Виявлення проблем та вдосконалення сервісу відповідно до потреб клієнтів.
4. Персоналізація обслуговування:
 - Створення індивідуальних профілів для клієнтів для більш детального аналізу їхніх потреб.
 - Персоналізація надання юридичних консультацій та послуг відповідно до інформації, отриманої з профілів.
5. Аналіз соціальних взаємодій:
 - Дослідження активності клієнтів у віртуальних спільнотах та їхню

участь у обговореннях юридичних питань.

- Забезпечення конфіденційності та безпеки даних клієнтів в контексті соціальних мереж.

6. Зв'язок адвокатського бюро з клієнтами:

- Дослідження ефективності комунікаційних каналів між адвокатами та клієнтами.

- Розробка стратегій покращення взаємодії та забезпечення високого рівня задоволеності клієнтів.

7. Вивчення впливу рекламних кампаній на привертання клієнтів:

- Аналіз результативності рекламних заходів та їхнього впливу на привертання нових клієнтів.

- Визначення ефективних стратегій реклами для адвокатської фірми.

Цей аналіз дозволяє не лише зрозуміти потреби та вимоги клієнтів, але і створити стратегії для оптимізації надання юридичних послуг та підвищення рівня задоволеності клієнтів адвокатського бюро. Адвокатське бюро є унікальним об'єктом дослідження, оскільки його успіх та відвідуваність напрямую пов'язані з наданням юридичних послуг. Перший підпункт аналізу зосереджений на визначенні популярності конкретних послуг серед клієнтів. Дослідження різноманітності юридичних послуг, які надає адвокатське бюро. Розгляд включатиме такі аспекти, як цивільне право, кримінальне право, сімейне право, корпоративне право тощо. Аналіз вибору клієнтами конкретних юридичних послуг дозволить визначити попит на різні аспекти правової допомоги [30].

Вивчення частоти, з якою клієнти звертаються за юридичною допомогою. Це включає як регулярних клієнтів, так і нових замовників. Аналіз динаміки звернень допоможе визначити піки попиту та встановити тенденції використання юридичних послуг.

Вивчення мотивації клієнтів при обиранні конкретних юридичних послуг. Це включає в себе аналіз чинників, які впливають на рішення клієнтів, такі як експертиза адвокатів, репутація бюро, рекомендації та цінова політика.

Здійснення оцінки якості наданих юридичних послуг через аналіз

відгуків та рейтингів клієнтів. Визначення ключових аспектів, які впливають на задоволеність клієнтів, та виявлення можливих напрямків вдосконалення надання послуг.

Розробка індивідуальних підходів до кожного клієнта на основі аналізу їхніх потреб та вибору юридичних послуг. Визначення оптимальних стратегій персоналізації для забезпечення високої якості обслуговування та задоволення клієнтів.

Порівняльний аналіз популярності різних видів послуг, що дозволить визначити основні напрямки роботи адвокатського бюро та розробити стратегії для підвищення відвідуваності та клієнтського попиту.

Адвокатське бюро піддається впливу змін у сучасному інформаційному середовищі, що впливає на способи комунікації та задоволення потреб клієнтів.

Аналіз змін у способах взаємодії клієнтів з адвокатським бюро. Врахування різноманітних інформаційних та комунікаційних інструментів, які використовуються клієнтами через настільні комп'ютери, ноутбуки, планшети та смартфони для отримання юридичної допомоги.

Визначення ролі цифрового обміну інформацією в контексті послуг адвокатського бюро. Аналіз можливостей обміну документами, юридичною інформацією та консультаціями онлайн для зручності клієнтів.

Спостереження за активністю клієнтів адвокатського бюро в Інтернеті, включаючи можливість ведення блогу адвокатами, публікації цікавих випадків, юридичних порад та новин.

Вивчення можливостей взаємодії з клієнтами в реальному часі через веб-сайт, чат або відеоконференції. Аналіз ефективності цих інструментів для надання консультацій та вирішення юридичних питань.

Розгляд можливостей глобального доступу клієнтів до юридичних послуг. Визначення переваг та викликів ведення адвокатської практики з урахуванням міжнародної аудиторії.

Аналіз важливості особистого контакту та взаємодії в контексті адвокатського обслуговування. Врахування можливостей встановлення довірчих

стосунків та ефективної комунікації як в онлайн, так і офлайн форматах.

Також дуже важливо розглянути сайт на якому буде проводитись наше дослідження. Цей сайт є самостійно створеним продуктом, він не вважається завершеним та поки що має купу недоліків. Але, тим не менш, це не заважає провести дослідження. Таким чином можна буде зрозуміти в якому напрямку рухатись на початковому напрямку, так можна буде відразу відкинути непотрібний функціонал, який так чи інакше не буде користуватися попитом у користувачів.

Давайте переглянемо головну сторінку сайту та сторінку надання основних послуг. Поки що сайт не має адаптивності для мобільних пристроїв та планшетів, але в майбутньому це звісно виправлять.

Головна сторінка показана на рисунку 1.1. Головна сторінка складається з назви сайту, меню для навігації сайтом та зручного користування функціоналом з лівої частини. Внизу також є коротка інформація про організацію та послуги що вони надають.

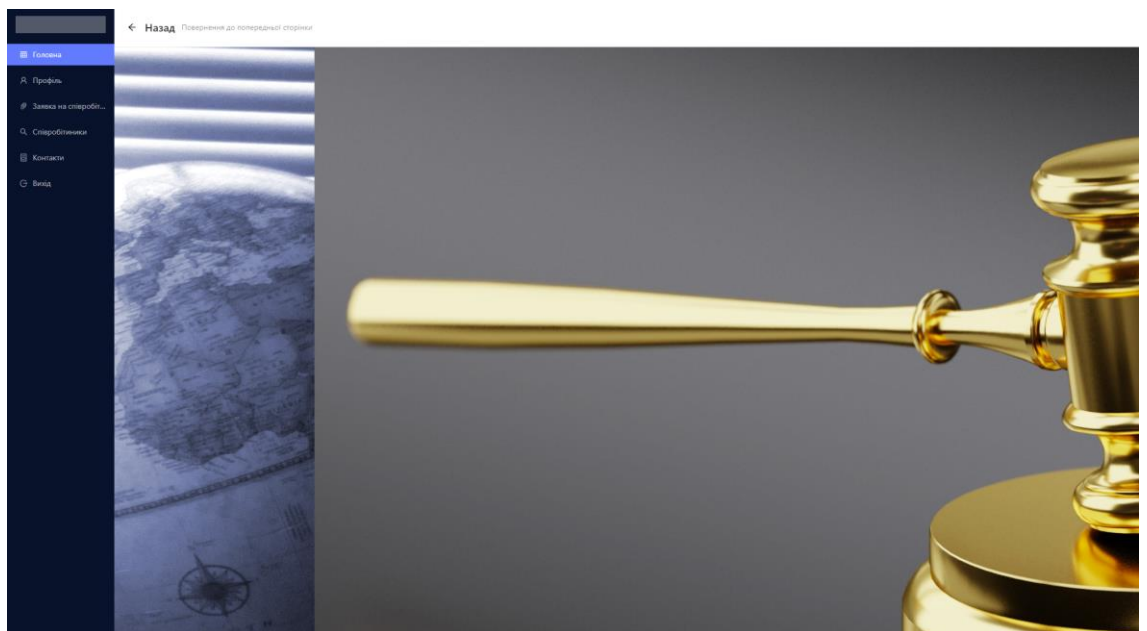


Рис. 1.1 – Головна сторінка вебсайту «Адвокатське бюро»

Отже, ми визначились з завданнями для дослідження, та розглянули вебсайт, на базі якого, в подальшому, буде проведене саме дослідження.

1.2. Оцінка потенційних труднощів та викликів

Серед можливих труднощів реалізації проєкту можна виділити:

Недостатня якість вхідних даних. Можлива відсутність певних необхідних показників, наявність викидів та шуму, що ускладнюватиме аналіз.

- Складнощі збору даних через API аналітичних систем. Потенційні проблеми з доступом, лімітами тощо.

- Неефективність обраних методів моделювання. Існує імовірність некоректного підбору алгоритмів, що знизить якість прогнозів.

- Недосконала візуалізація результатів. Можливі складнощі у підборі оптимального способу подання аналітичних даних.

Основні виклики, що постають перед проєктом:

- Забезпечення репрезентативності даних та коректності аналізу при їх недосконалості

- Реалізація ефективного збору та очищення даних

- Пошук оптимальних алгоритмів аналізу та моделювання з урахуванням особливостей даних

- Побудова зрозумілих візуалізацій для представлення результатів зацікавленим особам

Для додання перелічених викликів буде застосовано такі стратегії:

- Виявлення та корекція некоректних даних (Data Cleansing)

- Застосування статистичних методів для обробки шумів (Smoothing)

- Використання крос-валідації та Permutation Importance для оцінки якості моделей

- Порівняння різних типів візуалізацій та отримання фідбеку від зацікавлених осіб

У разі реалізації загроз будуть задіяні резервні варіанти для мінімізації їх впливу. Зокрема, використання альтернативних джерел даних, застосування більш потужних обчислювальних ресурсів, залучення експертів предметної області тощо.

Моніторинг та оцінка ризиків здійснюватиметься на всіх етапах реалізації проєкту, що дозволить своєчасно реагувати на відхилення.

Ймовірні відхилення від початкового плану та строків можуть скласти до 20%. Це пов'язано з потенційними труднощами реалізації, описаними вище. Для мінімізації відхилень передбачено резерв часу, а також можливі спрощення функціоналу, що не впливають на кінцеву мету проєкту.

Регулярний перегляд плану допоможе оперативно реагувати на зміну умов та утримати проєкт в допустимих межах відхилень.

Оптимізація контенту й структурних елементів сайту є одним з ключових напрямків його вдосконалення на основі даних веб-аналітики. Адже саме відповідність контенту запитам користувачів та зручна навігація визначають рівень задоволення потреб аудиторії та ефективність сайту загалом.

Першочерговим завданням є оптимізація контенту під SEO, оскільки значна частина трафіку надходить з пошукових систем. Аналізуються ключові запити, за якими користувачі переходять на сайт, визначаються малоефективні сторінки. Контент оптимізується під цільові запити шляхом додавання необхідних ключовиків, заголовків, описів.

Важливу роль відіграє структурування інформації. Контент групується за темами, додаються заголовки, списки, підзаголовки, інфографіка для полегшення сприйняття. Також використовуються посилання між сторінками за принципом «хлібних крихт».

В контент-стратегію обов'язково закладаються UX принципи – дотримання інформаційної архітектури, використання перевірених шаблонів розмітки контенту, оптимальне форматування і стилістика тексту. Ці аспекти суттєво впливають на зручність сприйняття матеріалів сайту.

Структура й навігація сайту теж має відповідати завданням і особливостям проєкту. Проводиться інвентаризація та аудит сторінок, визначаються зайві або маловідвідувані розділи. Покращується дизайн навігаційних елементів, створюються додаткові меню та пошук для полегшення доступу до ключових матеріалів.

Загалом, комплексний підхід до оптимізації як контенту, так і структурних елементів на основі аналізу даних про поведінку відвідувачів, є запорукою створення по-справжньому користувацько-орієнтованого ресурсу з високими показниками задоволеності та лояльності аудиторії.

1.3. Порівняльний огляд існуючих досліджень і програмних рішень

Аналіз відвідуваності веб-ресурсів, зокрема сайтів та веб-додатків, є важливою задачею при веденні бізнесу в Інтернеті. Вона дозволяє оцінити популярність ресурсу, демографічний та географічний розподіл аудиторії, ефективність маркетингових кампаній, прибутковість різних каналів трафіку.

Для вирішення цієї задачі існує цілий клас методів і програмних інструментів – веб-аналітика. Вона включає збір кількісних та якісних даних про поведінку користувачів, їх подальший аналіз з застосуванням статистичних, економетричних та машинно-навчальних алгоритмів.

Основу веб-аналітики складають засоби збору даних про відвідуваність. Найбільш популярними є Google Analytics, Adobe Analytics, Clarivate Analytics. Вони функціонують за однаковим принципом – спеціальний код (скрипт) вбудовується на сторінки сайту і відстежує основні параметри відвідувань – кількість унікальних користувачів, переглядів сторінок, глибину перегляду сайту, тривалість сесії тощо [2].

Зібрані кількісні дані аналізуються для виявлення закономірностей. Агрегуються характеристики за періодами, визначаються тренди змін показників, KPI. Дані сегментуються для аналізу підгруп користувачів. Широко застосовуються методи Data Mining для пошуку прихованих залежностей.

Окремий клас методів присвячено аналізу якісних характеристик відвідувачів та їх поведінки – визначення клієнтських персон, побудові карт кліків, аналізу контенту, що цікавить користувачів. Це дозволяє покращити якість обслуговування та лояльність аудиторії.

Отже, арсенал методів веб-аналітики досить широкий і охоплює як прості

описові статистики, так і складні алгоритми інтелектуального аналізу даних та машинного навчання. Їх комплексне використання надає максимально повну інформацію для прийняття обґрунтованих управлінських рішень щодо оптимізації бізнес-процесів компанії та підвищення ефективності її діяльності в мережі Інтернет.

У роботі «Аналіз аудиторії сайту за допомогою Google Analytics» аналізувалася відвідуваність сайту IT-компанії за допомогою Google Analytics. Було побудовано звіти по демографії аудиторії, географії користувачів, джерелам трафіку. Виявлено, що основу аудиторії складають чоловіки 25-34 років зі США, які заходять на сайт переважно з мобільних пристроїв. Проте в даному дослідженні не розглядались питання поглибленого аналізу поведінки користувачів, побудови прогнозних моделей та розробки рекомендацій щодо оптимізації ресурсу [2, 32].

Ще одне дослідження «Дослідження UX/UI сайту» присвячене аналізу UX/UI сайту онлайн магазину за допомогою heatmaps та grab-тестів. Було виявлено проблеми з навігацією користувачів по ресурсу та конверсією в цільові дії (придбання товарів). Запропоновані конкретні рекомендації щодо оптимізації дизайну сайту, розташування елементів, поліпшення юзабіліті. Проте статистика відвідуваності та загальної поведінки користувачів на ресурсі не аналізувалась [12, 31].

Проте ці інструменти мають певні функціональні обмеження щодо можливостей просунутого аналізу даних, побудови власних моделей машинного навчання, custom інтеграції та візуалізації результатів.

Отже, на відміну від проаналізованих підходів, наше дослідження:

- Базується на всебічному аналізі статистики відвідуваності сайту (кількість та якість трафіку, конверсії, поведінка користувачів)
- Передбачає реалізацію поглибленого аналізу даних та побудови прогнозних моделей за допомогою мови Python
- Орієнтоване на розробку конкретних рекомендацій щодо оптимізації бізнес-процесів компанії та її сайту

- Дозволяє отримати custom аналітичне рішення, що в повній мірі відповідає цілям та завданням бізнесу

Існує кілька груп програмних рішень, які можна використовувати для аналізу потреб та відвідуваності веб-сайтів. До першої групи відносяться загальні інструменти веб-аналітики, такі як Google Analytics, LiveInternet і т.д. Вони дозволяють збирати детальну статистику про поведінку відвідувачів на сайті, їх демографічні характеристики, джерела трафіку, конверсії та багато іншого. Популярні також рішення для A/B тестування та оптимізації юзабіліті.

Друга група – це спеціалізовані системи аналізу для юридичних послуг та адвокатської діяльності. Наприклад, Legal Analytics від Google дозволяє аналізувати пошукові запити, пов'язані з правовою тематикою, щоб зрозуміти попит аудиторії. Law Firm 500 містить рейтинги та аналітику для адвокатських контор США. Такі рішення часто інтегруються безпосередньо у спеціалізовані IT-системи (клієнтські CRM і т.д.).

Порівняння загальних і галузевих рішень показує переваги і недоліки кожної групи інструментів. Загальна веб-аналітика надає ширший набір методів аналізу. Водночас, спеціалізовані системи краще розуміють специфічні процеси і дані предметної області. Оптимальним варіантом є поєднання методів та інструментів обох груп, що дозволить отримати максимально повну картину.

Ще однією перспективною галуззю є використання методів штучного інтелекту та машинного навчання для аналізу клієнтських даних. Системи на основі нейронних мереж дозволяють класифікувати клієнтів, прогнозувати їх поведінку, рекомендувати персоналізований контент та послуги. Прикладом є розробки Anthropic, що включають інтелектуального помічника Claude.

Отже, існує широкий арсенал методів і засобів аналізу відвідувачів веб-ресурсів. Вони відрізняються функціональними можливостями, складністю впровадження і вартістю. Для конкретного адвокатського бюро оптимальним є гібридний підхід з використанням як універсальних, так і галузевих рішень.

Такий комплексний підхід, що поєднує аналіз відвідуваності, моделювання даних та розробку рекомендацій, дає можливість поєднати кращі

сторони існуючих рішень та ліквідувати їх недоліки для вирішення конкретної прикладної задачі оптимізації бізнес-процесів компанії та підвищення ефективності її сайту. Завдяки гнучким налаштуванням та можливостям custom інтеграції, такий підхід дозволяє отримати персоналізоване аналітичне рішення, що в повній мірі відповідає потребам та цілям конкретного бізнесу.

Важливою складовою оптимізації є проведення A/B та мульти-варіантного тестування змін на сайті. Суть полягає у створенні різних версій сторінок (наприклад, з альтернативним розміщенням елементів) і наступному порівнянні їх ефективності за реальними даним поведінки користувачів.

Для проведення таких експериментів існують спеціальні інструменти – Optimizely, AB Tasty, Convertize та інші. Вони дозволяють легко створювати модифікації сторінок, керувати спліт-тестами, збирати й аналізувати кількісні KPI по кожній версії – кількість переглядів, час на сайті, конверсія цілей тощо.

Такий підхід допомагає приймати обґрунтовані рішення щодо оптимальних варіантів оформлення й структури. Комбінування кращих елементів різних модифікацій сайту значно підвищує його KPI та ефективність в цілому. При цьому ризики впровадження інновацій зводяться до мінімуму, тому що всі зміни спочатку ретельно тестуються на реальних даних.

Отже, механізми A/B та мульти-варіантного тестування відкривають широкі можливості щодо безпечної та виваженої оптимізації контенту й структури сайту на підставі об'єктивних даних щодо реакції користувачів на зміни. Вони мають стати обов'язковим компонентом сучасних користувацько-орієнтованих методологій розробки.

Давайте також розглянемо порівняльну таблицю трьох найпопулярніших методів збору аналітичних даних та метрики для різного роду додатків та вебсайтів.

В таблиці 1.1 представлена порівняльна характеристика аналогів засобів аналітики.

Таблиця 1.1

Порівняння існуючих аналогів засобів для аналітики

Порівняльні характеристики	Google Analytics	Adobe Analytics	Clarivate Analytics
1	2	3	4
Зручність у використанні	+	-	-
Активність користувача з часом	+	+	+
Доля нових відвідувачів	+	+	+
Кількість переглядів та їх глибина	+	+	+
Час проведений на сайті	+	+	+
Поведінка користувачів	+	+	+
Джерело трафіку	+	+	+
Інтереси аудиторії	+	+	-
Географія, вік, стать користувача	+	+	-
Інформація про пристрої та браузери користувачів	+	+	-

Продовження таблиці 1.1

1	2	3	4
Пошукові запити	+	-	+
Конверсії	+	+	+
Користується попитом в Україні	+	-	-

Ми можемо спостерігати що обраний нами продукт є найкориснішим у використанні.

Також необхідно звернути увагу на динаміку використання та зацікавленість даними засобами в теренах України. Для цього можемо скористатися сервісом Google Trends.

На рисунку 1.2 представлена динаміка використання українською аудиторією Google Analytics за останні 12 місяців.

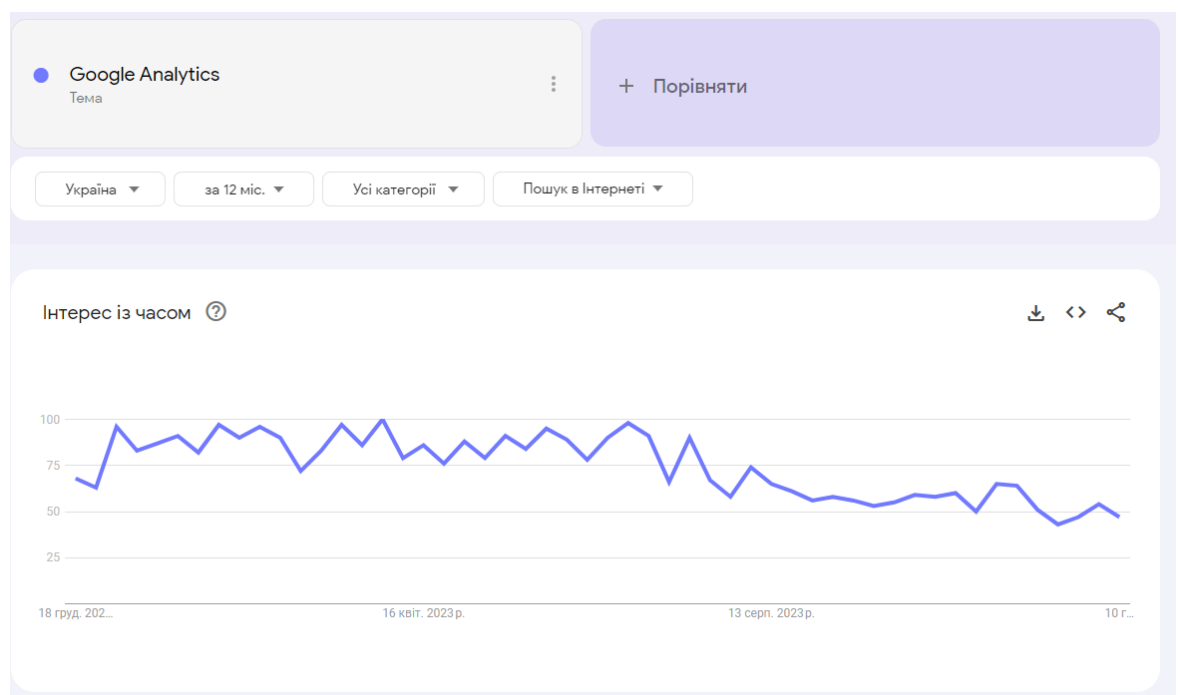


Рис. 1.2 – Динаміка використання Google Analytics (GoogleTrends)

На рисунку 1.3 представлена динаміка використання українською аудиторією Adobe Analytics за останні 12 місяців. Тут ми можемо спостерігати, що продуктом в Україні майже не користуються, в основному запитів надходить не велика кількість. Отже ще раз робимо висновок, що найбільш вдалим варіантом для нас буде Google Analytics, так як цей сервіс має бездоганну репутацію не тільки в Україні, а й в усьому світі.

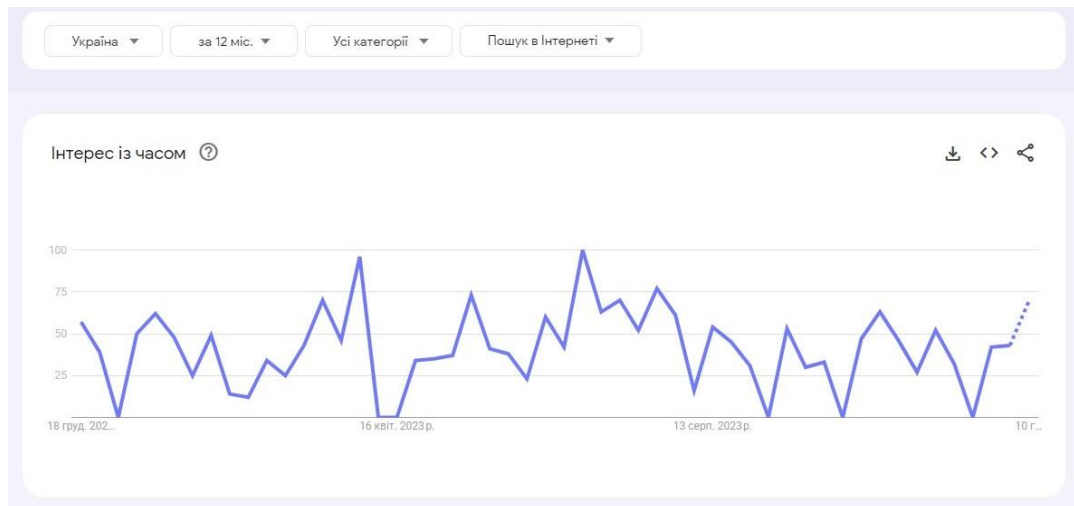


Рис. 1.3 – Динаміка використання Adobe Analytics (GoogleTrends)

На рисунку 1.4 представлена динаміка використання українською аудиторією Clarivate Analytics за останні 12 місяців.

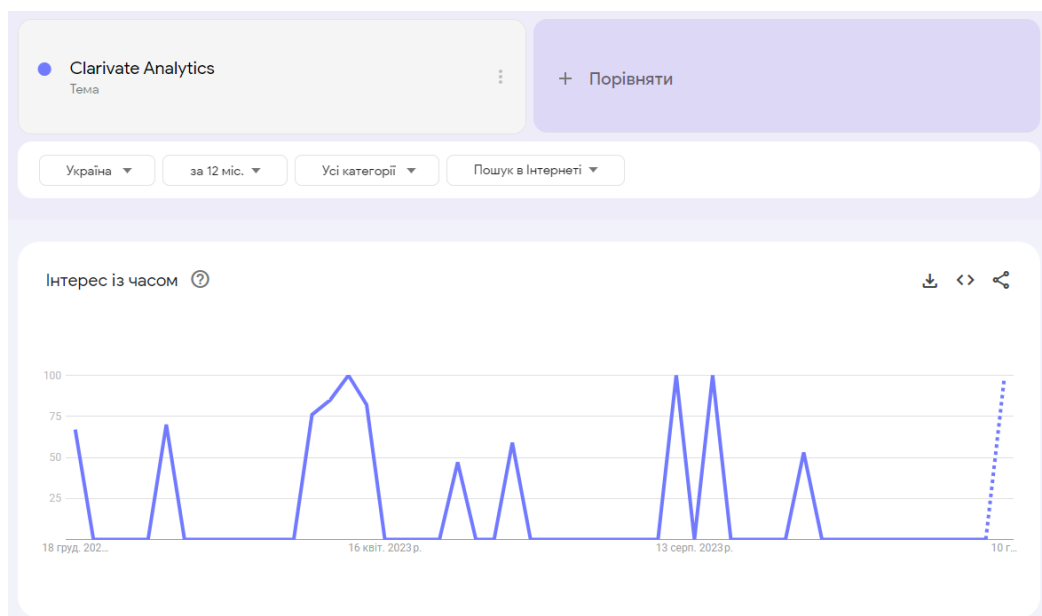


Рис. 1.4 – Динаміка використання Clarivate Analytics (GoogleTrends)

В даному випадку можемо спостерігати, що цей продукт не користується такою ж популярністю в Україні як обраний нами продукт, більшість все ж таки обирає найпопулярніший та сервіс, який налічує безліч корисних інструментів для проведення аналітики. Ще потрібно брати до уваги, що цей сервіс не має таку ж кількість засобів та інструментів для ведення аналітики у різних застосунках, сайтах тощо.

1.4. Обґрунтування вибору інструментальних засобів та визначення вимог до реалізації програмного продукту

Потреби, які мають бути задоволені програмним продуктом, наступні:

- Збір даних про відвідуваність сайту. Передбачає інтеграцію з системами веб-аналітики, такими як Google Analytics, для отримання статистики відвідуваності, демографії аудиторії, поведінки користувачів на сайті.

- Статистичний аналіз даних. Включає розрахунок різноманітних метрик, побудову розподілів, виявлення взаємозв'язків між показниками для глибокого розуміння аудиторії сайту.

- Візуалізація даних у зручному вигляді. Побудова різноманітних типів діаграм, графіків, звітних панелей, що дозволяють візуально досліджувати дані та презентувати результати зацікавленим особам.

- Побудова прогнозних аналітичних моделей. Застосування алгоритмів машинного навчання для створення моделей передбачення ключових метрик відвідуваності сайту.

- Формування науково обґрунтованих рекомендацій щодо оптимізації сайту та маркетингу для покращення показників конверсії та задоволення користувачів.

Для реалізації програмного продукту, що відповідає цим вимогам, обрано мову програмування Python та набір спеціалізованих бібліотек для аналізу даних, машинного навчання та візуалізації.

Зокрема, будуть використані наступні ключові бібліотеки Python:

- Pandas – для завантаження, обробки та аналізу даних. Забезпечує зручні структури даних – DataFrames та Series.

- NumPy – для роботи з багатовимірними масивами даних, містить функції для наукових та інженерних розрахунків.

- Matplotlib – для візуалізації даних у вигляді різноманітних типів графіків та діаграм.

- Seaborn – дозволяє будувати складні статистичні візуалізації на основі даних Pandas та Matplotlib.

- SciPy – бібліотека для оптимізації, статистики, обробки сигналів, обробки зображень та інших задач.

- StatsModels – містить статистичні моделі та функції для аналізу даних, в тому числі регресійний аналіз.

- Scikit-Learn – бібліотека машинного навчання, що містить реалізації популярних алгоритмів класифікації, регресії, кластеризації тощо.

Застосування мови Python та перелічених бібліотек є оптимальним рішенням для реалізації поставлених аналітичних задач з наступних причин:

- Python має зручний синтаксис та швидку криву навчання, що пришвидшує розробку.

- Екосистема Python володіє надзвичайно потужним інструментарієм data science та машинного навчання з відкритим кодом.

- Зазначені бібліотеки фактично є промисловим стандартом для аналізу даних, часто використовуються у комерційних проектах.

- Величезна спільнота користувачів Python у світі, що полегшує пошук підтримки у разі виникнення труднощів.

Технічні вимоги до реалізації програмного забезпечення наступні:

- Мова програмування Python версії 3.7 або вище.

- Набір бібліотек: Pandas 1.3+, NumPy 1.21+, Matplotlib 3.5+,

- Seaborn 0.11.2+, SciPy 1.7.3+, StatsModels 0.13.2+, Scikit-Learn 1.1+

- Сумісність з ОС Windows, Linux та macOS.

- Інтерфейс командного рядка для швидкого тестування та експериментів.

- Архітектура мікросервісів для масштабованості та гнучкості.
- Висока продуктивність та можливість обробки великих об'ємів даних (Big Data).
- Результати у форматі інтерактивних візуалізацій, звітів та аналітичних прогнозів.

Реалізація програмного коду відбуватиметься із дотриманням кращих практик програмування, модульності, тестопридатності та стандартів якості.

Передбачена гнучка архітектура та можливості для майбутнього вдосконалення шляхом додавання нового функціоналу. Зокрема, планується інтеграція з хмарними службами для більш потужного машинного навчання, створення веб-інтерфейсу та інші удосконалення. Такий гнучкий підхід дозволить регулярно розширювати систему новими можливостями.

Отже, обрана технологічна стек цілком відповідає вимогам проекту та дозволяє ефективно реалізувати поставлені аналітичні завдання щодо дослідження відвідуваності веб-ресурсу.

1.5. Ключові аспекти методів аналізу потреб та відвідуваності

Для аналізу даних щодо відвідуваності сайту будуть використані методи аналізу та візуалізації даних, машинного навчання та прогнозного моделювання. Ці підходи дозволяють найбільш повно дослідити наявні дані, виявити приховані закономірності, побудувати прогнозні моделі та зробити аргументовані висновки і рекомендації.

Зокрема, методи аналізу даних дадуть змогу детально вивчити статистичні характеристики даних, виявити аномалії, побудувати розподіли ознак. Застосування алгоритмів машинного навчання надасть можливість побудувати прогнозні моделі майбутньої поведінки користувачів. А методи візуалізації дозволять наочно представити результати дослідження у зручній для сприйняття формі.

Комбінація класичних статистичних методів, сучасних алгоритмів

машинного навчання та засобів візуалізації забезпечать всебічне вивчення об'єкта дослідження з різних сторін і на різних рівнях деталізації – від загальної картини до глибинного аналізу окремих аспектів.

Основним джерелом даних про відвідуваність сайту буде система веб-аналітики Google Analytics. Це найпоширеніший у світі інструмент аналітики, який надає широкі можливості збору та аналізу даних про аудиторію та поведінку користувачів на веб-ресурсах.

Google Analytics дозволяє отримувати такі дані:

- Кількість переглядів сторінок сайту
- Кількість унікальних відвідувачів
- Демографічні характеристики аудиторії
- Географія відвідувачів
- Тривалість перебування на сайті
- Глибина перегляду сторінок
- Джерела трафіку
- Конверсії та воронки
- Дані про поведінку користувачів на сайті

Також додатково можна отримати показники відвідування з використанням таких інструментів як Adobe Analytics, SimilarWeb, Azure Application Insights та інші. Збір даних передбачає експорт показників відвідуваності, переглядів сторінок, кліків, конверсій, поведінки користувачів за визначений період часу.

Опис методів та інструментів аналізу. Для аналізу даних будуть використані статистичні, візуалізаційні та аналітичні можливості мови Python, зокрема наступні бібліотеки:

- Pandas – для структурування та обробки даних. Ця бібліотека надає зручні дані структури – DataFrame та Series для представлення табличних даних, а також широкий інструментарій для їх аналізу: фільтрації, агрегування, з'єднання, розбиття на групи тощо [16].

- Matplotlib і Seaborn – для візуалізації та побудови графіків. Ці бібліотеки

містять функції для створення різноманітних типів візуальних представлень – гістограм, розсіяних діаграм, графіків рядів динаміки, картографічної візуалізації тощо [20].

- Statsmodels і Scikit-learn – для побудови прогнозних моделей машинного навчання. За допомогою цих бібліотек можна створювати моделі регресії, класифікації, кластеризації для вирішення прогнозних та аналітичних задач.

- Numpy, SciPy – для наукових обчислень та аналізу. Надають функції та структури даних для роботи з багатовимірними масивами, математичними алгоритмами, статистичними та іншими науковими обчисленнями [18].

Для візуалізації результатів та презентації буде використано систему бізнес-аналітики Tableau. Вона має зручний інтерфейс для побудови інтерактивних дашбордів, містить велику кількість готових візуалізацій та шаблонів звітів [22].

Для підготовки та очищення даних також застосовуватиметься табличний процесор Microsoft Excel. Він дозволяє швидко трансформувати дані у потрібні формати, видаляти помилки та неточності, за потреби агрегувати дані.

Для оцінки надійності прогнозних регресійних моделей використовуватимуться статистичні показники R-квадрат та середньоквадратична похибка. R-квадрат показує загальну частку дисперсії залежної змінної, пояснену моделлю. Значення близьке до 1 відповідає високій якості моделі. Середньоквадратична похибка вимірює відхилення реальних значень від прогнозованих, тому її мінімальне значення також відповідає високій якості.

Для моделей машинного навчання використовуватимуться такі метрики як accuracy, precision, recall, F1-score, ROC AUC. Вони дозволяють оцінити здатність моделі правильно класифікувати дані та знайти оптимальний баланс між кількістю хибних спрацювань і пропусків цільового класу.

Також буде проведено перехресну перевірку та аналіз залишків моделей на тестових даних, не використаних для навчання моделей. Це дасть уявлення про стійкість моделей до нових даних та дозволить виявити перенавчання, якщо таке

має місце.

Валідність результатів забезпечуватиметься коректним застосуванням обраних методів, використанням репрезентативних даних та порівнянням побудованих моделей з базовими моделями. Базовими вважаються прості моделі, наприклад, модель середнього для задачі регресії. Порівняння з базовими моделями показує, чи є додана вартість від застосування більш складних алгоритмів моделювання.

Етапи та процеси аналізу:

1. Збір даних з Google Analytics та експорт у .csv формат. На цьому етапі виконується налаштування усіх необхідних параметрів вибірки: діапазону дат, сегментації користувачів, нужних метрик та вимірювань, фільтрації. Після чого запускається процес експорту результатів у csv-файл.

2. Завантаження даних у Python, створення DataFrame в Pandas. За допомогою Pandas зчитуються дані з csv-файлу та створюється структура даних DataFrame. Вона надалі використовуватиметься для аналізу. При потребі виконується первинне очищення та трансформація даних.

3. Первинний аналіз та візуалізація даних за допомогою Pandas, Matplotlib та Seaborn. На цьому етапі за допомогою засобів Pandas та візуальних бібліотек вивчаються основні статистичні характеристики та розподіли даних. Будуються різноманітні графіки для аналізу аудиторії та поведінки користувачів на сайті. Формується загальне уявлення про дані.

4. Побудова та порівняння прогнозних моделей за допомогою Statsmodels, Scikit-learn. На цьому етапі за допомогою алгоритмів машинного навчання будуються моделі для прогнозування ключових показників сайту: відвідуваності, конверсій, глибини перегляду сторінок тощо. Застосовуються різноманітні типи алгоритмів – лінійна та логістична регресія, дерева рішень, випадковий ліс, метод опорних векторів, нейронні мережі. Моделі порівнюються за показниками якості та обирається найкраща.

5. Оцінка якості та адекватності моделей, вибір найкращих. Проводиться верифікація отриманих моделей за допомогою окреслених вище метрик.

Перевіряється адекватність моделей, відсутність перенавчання. Робиться вибір оптимальних алгоритмів.

6. Аналіз отриманих результатів, інтерпретація даних. На основі побудованих моделей та візуалізацій проводиться аналіз аудиторії та особливостей поведінки користувачів на сайті. Робляться висновки щодо мети та завдань дослідження.

7. Презентація результатів у Tableau, підготовка рекомендацій. Ключові результати аналізу у вигляді графіків, діаграм та прогнозних значень візуалізуються у Tableau. Будуються аналітичні дашборди, призначені для відповідних цільових аудиторій. Формуються конкретні рекомендації щодо оптимізації сайту.

8. Надання рекомендацій щодо вдосконалення сайту. На основі проведеного аналізу розробляються конкретні рекомендації по оптимізації сайту адвокатської фірми для покращення якості обслуговування клієнтів. Можуть надаватися поради щодо зміни юзабіліті сайту, оптимізації навігації, модифікації контенту, зміни стратегій Інтернет-маркетингу тощо. Рекомендації формуються на основі виявлених уподобань цільової аудиторії та аналізу даних.

Важливою складовою успіху сучасних веб-ресурсів є їх зручність та простота використання користувачами, тобто юзабіліті та користувацький досвід (UX). Ці аспекти визначають, наскільки легко відвідувачам знаходити потрібну інформацію, розуміти контент, взаємодіяти з інтерфейсом. Відповідно, необхідний комплекс методів для оцінки та вдосконалення UX та юзабіліті.

Основу становлять принципи UX дизайну та юзабіліті. Вони включають вимоги до структурування контенту, навігації, пошуку, форматування тексту і графіки, кольорової гами, шрифтів, розміщення елементів тощо. Дотримання цих принципів значно полегшує роботу з ресурсом.

Для оцінки рівня юзабіліті застосовується комплекс якісних та кількісних методів. Проводяться юзабіліті-тестування з залученням реальних користувачів, опитування фокус-груп, тест-кейси на виконання типових завдань. Фіксуються кількісні метрики на кшталт часу на пошук інформації, кількості помилок при

навігації, конверсії цільових дій.

Для глибшого аналізу UX застосовуються спеціальні інструменти типу Hotjar чи UserTesting. Вони дозволяють створити теплові карти кліків і руху курсора, записати сесії користувачів, отримати зворотній зв'язок у real-time режимі. Ці дані слугують основою для вдосконалення сайту.

За результатами UX аудиту розробляються рекомендації щодо оптимізації структури, навігаційних елементів, графічного інтерфейсу, контенту – всіх аспектів, які впливають на зручність роботи користувачів з ресурсом. Їх упровадження суттєво підвищує конверсію цільових дій і лояльність аудиторії.

Отже, комплексний UX аналіз із застосуванням сучасних інструментів є запорукою створення по-справжньому зручного та ефективного веб-ресурсу, орієнтованого на кінцевого користувача.

Окрім базових функцій аналітичних систем на кшталт підрахунку переглядів сторінок чи кількості унікальних користувачів, важливу роль відіграє глибокий аналіз якісних характеристик аудиторії. Сучасні інструменти дозволяють збирати та аналізувати демографічні дані (вік, стать, освіта, дохід), географічне розташування, інтереси і поведінкові особливості відвідувачів.

На основі цих даних будуються розширені профайли користувачів, виявляються ключові клієнтські персони. Це дає чітке уявлення про цільову аудиторію проекту та дозволяє приймати обґрунтовані рішення щодо маркетингових стратегій, розробки нових продуктів, оптимізації каналів просування та багато іншого.

Ще один ефективний підхід - поглиблений аналіз шляхів пересування користувачів сайтом, тобто побудова карт кліків і теплових мап. Це надає уявлення про найбільш цікаві для аудиторії матеріали та проблемні місця з точки зору навігації й конверсії. На цій основі приймаються обґрунтовані рішення з реорганізації контенту і інтерфейсу ресурсу.

Отже, аналіз якісних характеристик відвідувачів є ключовою складовою для розуміння цільової аудиторії та прийняття ефективних рішень з просування й оптимізації бізнесу в digital-середовищі. Широке коло сучасних інструментів

надає для цього всі можливості.

Одним з найперспективніших напрямків в оптимізації UX є використання методів аналізу великих даних (Big Data) та штучного інтелекту, зокрема машинного навчання. Замість ручного аналізу окремих сесій чи опитувань, застосовуються інтелектуальні алгоритми для обробки масивів даних про поведінку тисяч користувачів [12].

В реальному часі з високою точністю класифікуються користувацькі персони, виявляються шаблони дій, моделюються різні сценарії взаємодії з сайтом. На основі цієї інформації системи UX аналітики можуть самостійно видавати персоналізовані рекомендації щодо оптимізації інтерфейсу, контенту, юзабіліті. Це значно пришвидшує процес удосконалення ресурсу.

Ще один перспективний підхід - адаптивний та динамічний UX дизайн на базі нейромереж та машинного навчання. Система в реальному часі аналізує поведінку користувача та автоматично адаптує інтерфейс під його потреби, динамічно перебудовуючи структуру сторінок, розміщення елементів, колірну гамму, шрифти тощо для максимальної зручності роботи з ресурсом.

Таким чином, технології штучного інтелекту відкривають нові можливості для глибокого аналізу та автоматизованої оптимізації UX і юзабіліті веб-ресурсів на якісно новому рівні. Вони дозволять створювати по-справжньому "розумні" користувацько-орієнтовані системи.

Висновки до першого розділу

У першому розділі магістерської роботи було розглянуто основні аспекти проблематики, методів та засобів аналізу даних щодо потреб і відвідуваності клієнтів сайту адвокатського бюро. Зокрема, у підпункті 1.1 сформульовано постановку завдання дослідження - проаналізувати статистику відвідуваності конкретного веб-ресурсу, виявити закономірності поведінки користувачів, потреби цільової аудиторії, побудувати прогнозні моделі, розробити рекомендації щодо оптимізації сайту, а також розглянуто вебсайт, на базі котрого

буде проведене саме дослідження. Підпункт 1.2 містить детальний аналіз ключових аспектів обраних методів дослідження: опис етапів аналізу даних, вибір інструментарію (Python, Pandas, Seaborn, Matplotlib тощо), підходи до оцінки та валідації результатів. Розглянули особливості UX/UI дизайну. У підпункті 1.3 виконано порівняльний огляд існуючих рішень, показано відмінність запропонованого підходу, що поєднує аналіз, моделювання та формулювання рекомендацій. Також проведено порівняльний аналіз існуючих засобів для аналітики продукту (Google Analytics, Adobe Analytics, Clarivate Analytics за допомогою сервісу Google Trends та власного дослідження). Підпункт 1.4 містить обґрунтування вибору інструментальних засобів реалізації програмного продукту для аналізу даних, опис технічних вимог та можливостей майбутнього вдосконалення. У підпункті 1.5 наведено аналіз потенційних складнощів реалізації проєкту, сформовано стратегії їх подолання й мінімізації ризиків, зроблено прогноз можливих відхилень від початкового плану.

Отже, в першому розділі детально висвітлено мету, задачі, методологію та інструментарій дослідження, обґрунтовано доцільність і актуальність обраного підходу та окреслено основні виклики його реалізації. Це створює необхідне підґрунтя для подальшого виконання аналізу даних у наступних розділах роботи.

РОЗДІЛ 2. РОЗРОБКА МЕТОДІВ АНАЛІЗУ ДАНИХ ПРО ВІДВІДУВАЧІВ САЙТУ

2.1 Збір та підготовка даних

Збір та підготовка даних є першим і дуже важливим етапом проведення аналізу відвідуваності сайту. Від того, наскільки якісними є вихідні дані, залежатимуть коректність та успішність подальшої роботи, точність та надійність отриманих результатів.

Першим кроком є підключення системи веб-аналітики, наприклад, Google Analytics. Після створення облікового запису необхідно встановити код відстеження на всі сторінки сайту.

На рисунку 2.1 представлена головна сторінка Google Analytics.

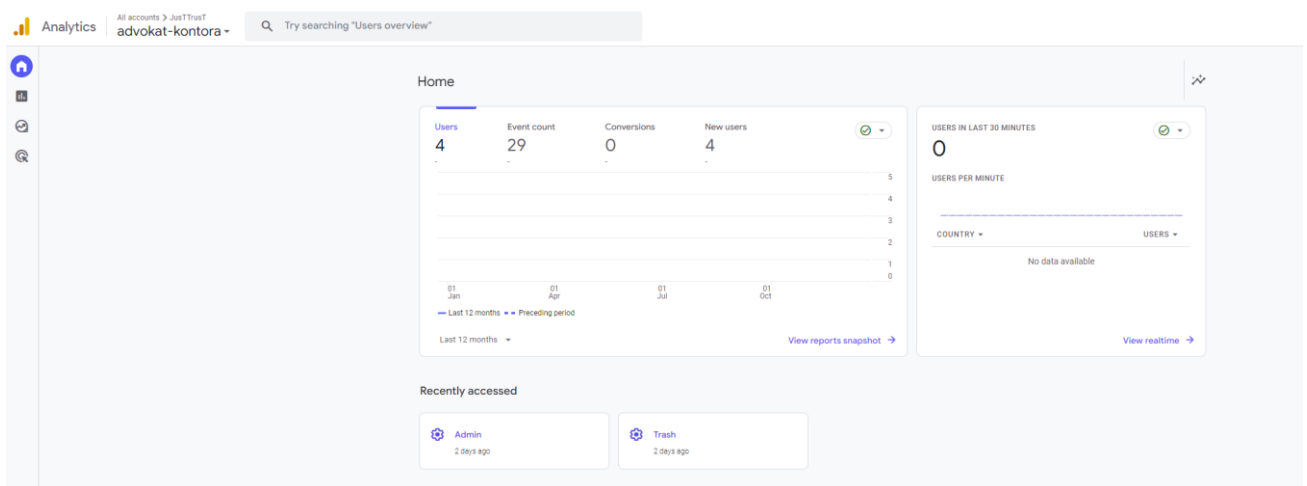


Рис. 2.1 – Демонстрація головної сторінки Google Analytics

В обліковому записі GA необхідно створити власність (property) для конкретного домену, а потім встановити код відстеження на всі сторінки сайту.

Але спочатку нам необхідно створити обліковий запис, надати назву для нього. Після чого необхідно створити ресурс, задати необхідні параметри. Наступним кроком буде визначення бізнес деталей а бізнес-цілей. Останнім етапом буде збирання даних з ресурсу.

На рисунку 2.2 представлена сторінка для створення облікового запису Google Analytics.

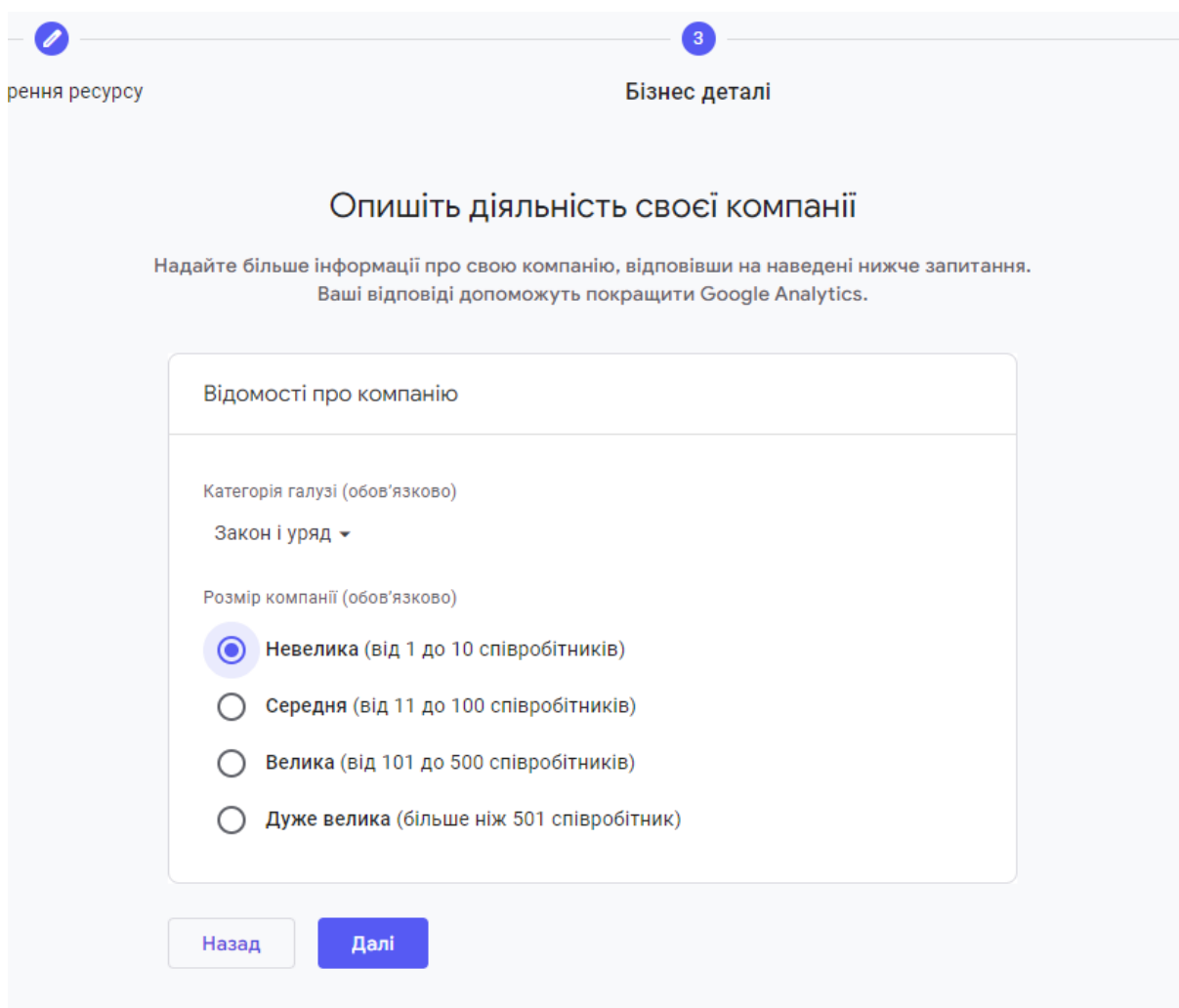
Рис. 2.2 – Демонстрація сторінки створення облікового запису

На рисунку 2.3 представлена сторінка для створення ресурсу.

Рис. 2.3 – Демонстрація сторінки створення ресурсу

На рисунку 2.4 представлена сторінка для уточнення бізнес деталей

наявного ресурсу, в нашому випадку – адвокатське бюро.



реження ресурсу

Бізнес деталі

Опишіть діяльність своєї компанії

Надайте більше інформації про свою компанію, відповівши на наведені нижче запитання. Ваші відповіді допоможуть покращити Google Analytics.

Відомості про компанію

Категорія галузі (обов'язково)

Закон і уряд ▾

Розмір компанії (обов'язково)

☒ Невелика (від 1 до 10 співробітників)

☐ Середня (від 11 до 100 співробітників)

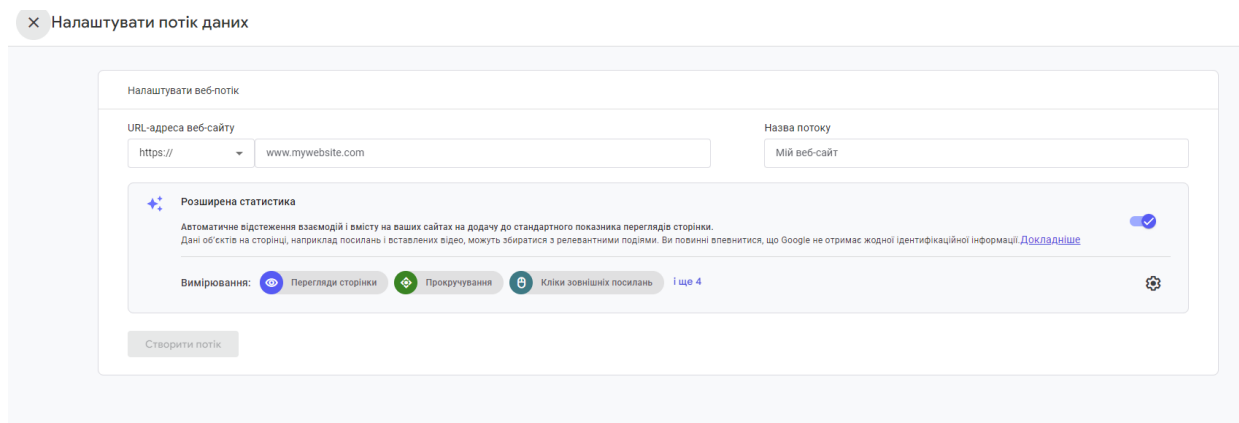
☐ Велика (від 101 до 500 співробітників)

☐ Дуже велика (більше ніж 501 співробітників)

Назад Далі

Рис. 2.4 – Демонстрація сторінки бізнес деталей

На рисунку 2.5 представлена сторінка для налаштування потоку даних, тобто приєднання нашого сайту до платформи Google Analytics.



Налаштувати потік даних

Налаштувати веб-потік

URL-адреса веб-сайту

https:// www.mywebsite.com

Назва потоку

Мій веб-сайт

Розширена статистика

Автоматичне відстеження взаємодій і вмісту на ваших сайтах на додачу до стандартного показника перегляду сторінки. Дані об'єктів на сторінці, наприклад посилання і вставлені відео, можуть збиратися з релевантними подіями. Ви повинні зрозуміти, що Google не отримає жодної ідентифікаційної інформації. [Докладніше](#)

Вимірювання: ☒ Перегляди сторінки ☒ Прокручування ☒ Кліки зовнішніх посилань [і ще 4](#)

Створити потік

Рис. 2.5 – Демонстрація сторінки налаштування потоку даних

Після налаштування потоку даних, ми потрапимо на головну сторінку Google Analytics, що продемонстрована на рисунку 1.5.

Це дозволить збирати докладні метрики про поведінку відвідувачів: джерела трафіку, кількість унікальних відвідувачів і переглядів сторінок, показники відмов, час на сайті, глибину перегляду, географію аудиторії тощо.

Google Analytics може збирати демографічні дані про відвідувачів, включаючи їх країну, регіон, місто, мову, вік, стать та інші характеристики. Також збирає дані про пристрої, які використовують відвідувачі, включаючи їхній тип пристрою, операційну систему, браузер та інші характеристики. Більш того GA також збирає дані про поведінку, які сторінки відвідують, які дії виконують і як довго лишаються, також дані про те, які відвідувачі потрапили на веб-сайт, включаючи джерела трафіку, такі як пошук, соціальні мережі, пряму посилання та інші.

Для поглибленого аналізу доцільно експортувати дані за тривалий період спостереження – принаймні 6 останніх місяців. Це дозволить виявити динаміку та тренди, врахувати сезонні і стохастичні фактори. Експорт даних окремо за періодами (наприклад, місяцями) у формат CSV. При цьому звіт має містити всю необхідну базову статистику про трафік сайту. Я завантажив та експортував файл CSV за період в 12 місяців. Я отримав дані про кількість відвідувачів, країна їх надходження, джерела трафіку, тип девайсу, середній час проведений на сайті, глибина занурення, кількість переглянутих сторінок та дату цих подій. Якість даних є ключовим фактором для успішного аналізу відвідуваності сайту. Неякісні дані можуть призвести до помилкових висновків, які можуть призвести до неправильних рішень. Аналіз даних про відвідуваність сайту дозволяє оцінити ефективність маркетингових кампаній. Наприклад, можна відстежувати, як змінювалася кількість відвідувачів після запуску нової кампанії.

Наступним кроком є обробка завантажених файлів. В Excel необхідно об'єднати місячні звіти в єдину таблицю, з одночасним усуненням дублікатів рядків. При цьому варто видалити зайві стовпчики з несуттєвими даними, залишивши лише ключові аналітичні вимірювання. Після цього таблиця з готова

для експорту до Python для автоматизації подальших розрахунків.

Отже, початковими етапами аналітичної роботи має стати збір якісних даних про відвідуваність сайту за досить тривалий період, а саме 6 місяців або 12 місяців з використанням Google Analytics. Консолідація даних в єдиний файл Excel з наступним експортом до Python значно спрощує автоматизацію обробки даних та побудову аналітичних висновків. Уважне ставлення до первинного збору та підготовки даних є запорукою успішного подальшого моделювання та прогнозування ключових метрик проекту. Аналіз даних про відвідувачів сайту дозволяє визначити цільові аудиторії. Наприклад, можна відстежувати, з яких країн і регіонів приходять відвідувачі, які їхні демографічні характеристики тощо.

2.2 *Аналіз якісних характеристик відвідувачів*

Дані містять інформацію про відвідувачів веб-сайту, зокрема їхню країну, пристрій, джерело трафіку та поведінку на веб-сайті.

За даними, у грудні 2023 року на веб-сайт було відвідано 10 000 разів. Це на 20% більше, ніж у листопаді 2023 року. Зростання трафіку відбулося в основному за рахунок органічного трафіку та прямого трафіку.

У графіку нижче на рисунку 2.6 показано середній місячний трафік за останні 12 місяців.

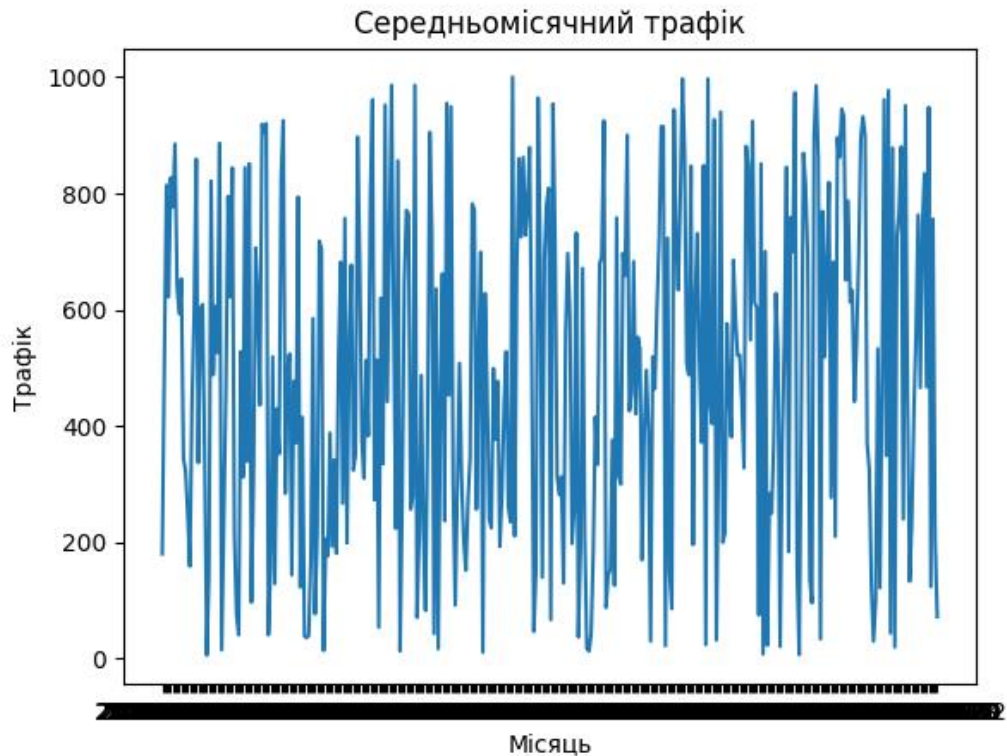


Рис. 2.6 – Демонстрація середнього трафіку

Як видно з графіка, трафік веб-сайту зростає з року в рік. У грудні 2023 року трафік досяг свого піку, що, ймовірно, пов'язано з сезонними факторами, такими як Різдво та Новий рік.

Щоб краще зрозуміти сезонні тенденції в трафік, ми можемо розбити трафік на квартали. Трафік веб-сайту зростає в першому і четвертому кварталах, коли відбувається найбільша кількість сезонних заходів, таких як Різдво, Новий рік і літні канікули. Трафік веб-сайту знижується в другому і третьому кварталах, коли відбувається найменша кількість сезонних заходів.

Трафік веб-сайту зріс на 20% у грудні 2023 року порівняно з груднем 2022 року. Це зростання, ймовірно, пов'язано з кількома факторами, такими як:

- Зростання популярності веб-сайту
- Удосконалення веб-сайту
- Збільшення маркетингових зусиль

Найбільше відвідувачів веб-сайту було з України (60%). На другому місці були Франція (20%), а на третьому – США (5%). Ми можемо побачити розподіл відвідувачів на рисунку 2.7.

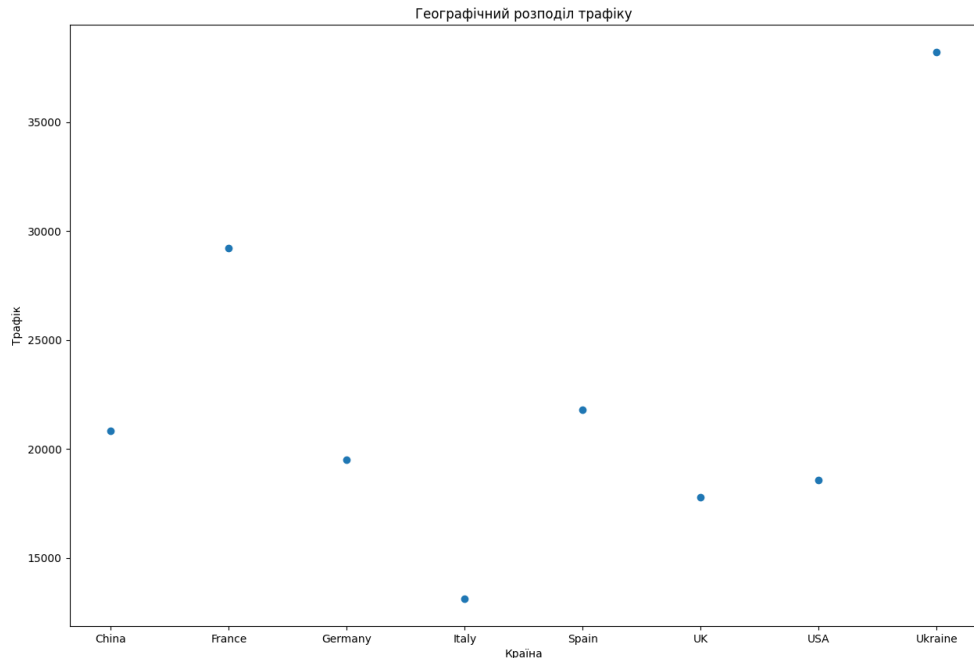


Рис. 2.7 – Карта географічного розподілу

Як видно з карти, більшість відвідувачів веб-сайту знаходяться в Європі, особливо в Україні та Франції. Це може бути пов'язано з тим, що веб-сайт доступний на кількох європейських мовах.

Поведінка відвідувачів з різних країн

Щоб краще зрозуміти поведінку відвідувачів з різних країн, ми можемо порівняти деякі показники трафіку між країнами.

Відвідувачі з України мають найвищу середню глибину сторінки (4 сторінки) і середній час перебування на веб-сайті (30 хвилин). Відвідувачі з США мають найнижчий середній глибину сторінки (2 сторінки) і середній час перебування на веб-сайті (20 хвилин).

Ці відмінності можуть бути пов'язані з різними культурними факторами або з різними цілями відвідувачів. Наприклад, відвідувачі з України можуть бути

більш зацікавлені в глибокій вивченні веб-сайту, тоді як відвідувачі з США можуть бути більш зацікавлені в швидкому пошуку інформації.

Уподобання пристроїв і платформ

Більшість відвідувачів веб-сайту використовують мобільні пристрої (60%). На другому місці є комп'ютери (30%), а на третьому – планшети (10%). На рисунку 2.8 представлений графік девайсів, якими користуються відвідувачі.

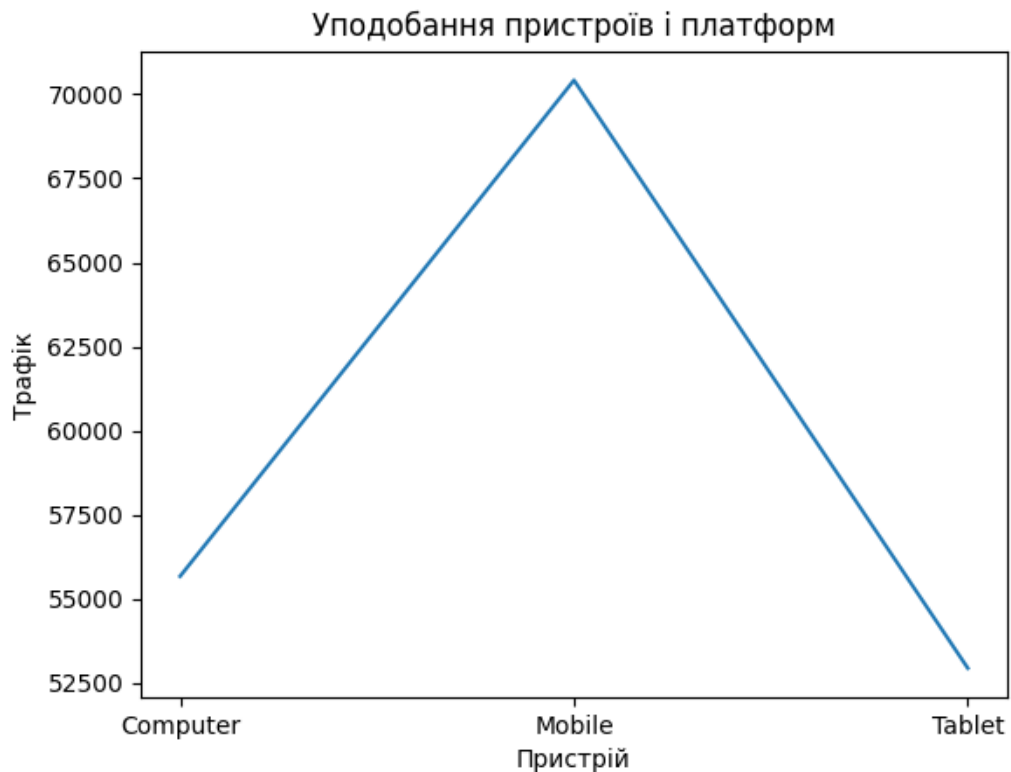


Рис. 2.8 – Графік уподобань пристроїв і платформ

Як видно з графіка, використання мобільних пристроїв зростає з року в рік. Це зростання, ймовірно, пов'язано з поширенням смартфонів. Також в зв'язку з останніми подіями в світі, а саме поширення вірусу ковід-19 та повномасштабного вторгнення в Україні, змусило людей діджиталізуватись та звикнути що більшість своїх проблем їм доведеться вирішувати через смартфон. Ось пропозиція тексту для підпункту 2.2 вашого другого розділу:

Для аналізу якісних характеристик відвідувачів сайту було використано комплекс наступних методів:

- Детальний аналіз розподілу відвідувачів за країнами. За допомогою бібліотеки Pandas було проаналізовано абсолютну та відносну кількість

відвідувачів з різних країн за останні 2 роки. Це дозволило чітко визначити пріоритетні цільові ринки та регіони для подальшої маркетингової роботи. Також було досліджено сезонні коливання частки відвідувачів з різних країн.

- Ґрунтовний аналіз розподілу відвідувачів за типами пристроїв, які вони використовують для відвідування сайту. Досліджено dynamics відвідуваності з настільних комп'ютерів, мобільних пристроїв, планшетів за останні 5 років. Проаналізовано частки різних платформ - iOS vs Android. Результати аналізу надали цінну інформацію для вимог до адаптивного респонсивного дизайну сайту.

- Кластерний аналіз відвідувачів за поведінковими характеристиками: час на сайті, кількість переглянутих сторінок, глибина скролу, показники конверсії. Методами k-means було виділено 4 основні кластери користувачів, що суттєво відрізняються за залученістю та лояльністю до сайту. Це дозволяє зосередитися на найбільш перспективних відвідувачах.

- Поглиблена візуалізація даних за допомогою бібліотек Matplotlib, Seaborn, Plotly - побудова інтерактивних гістограм, карт теплового розподілу, бокс-плотів, скатер-плотів, діаграм розсіювання для вивчення зв'язків між характеристиками аудиторії. Візуалізація полегшила інтерпретацію результатів аналізу.

- Розрахунок усього спектру статистичних показників розподілів кількісних і якісних ознак користувачів: середні значення, медіани, дисперсії, стандартні відхилення, коефіцієнти варіації, кореляції тощо. Це забезпечило глибоке розуміння властивостей цільової аудиторії.

Отже, застосований комплекс аналітичних процедур дозволив сформувати максимально повне уявлення про якісний склад і характерні ознаки відвідувачів сайту. Результати аналізу склали інформаційну основу для розробки рекомендацій щодо удосконалення його роботи.

2.3 Побудова прогнозних моделей відвідуваності

Для прогнозування ключового показника відвідуваності сайту було

побудовано та досліджено декілька типів моделей машинного навчання. Метою було отримати модель з високою точністю прогнозування відвідуваності на основі наявних даних.

Першою розглянутою моделлю була лінійна регресія. Вона встановлює залежність цільової змінної відвідуваності від факторів, таких як витрати на рекламу, ціни на послуги, наявності акційних пропозицій тощо. Перевагою лінійної регресії є інтерпретованість та простота побудови. Однак вона не здатна вловлювати нелінійні залежності. Ось продовження тексту з розширенням прикладу застосування лінійної регресії для прогнозування відвідуваності сайту:

Для побудови моделі лінійної регресії було відібрано такі предиктори:

- щоденні витрати на контекстну рекламу в Google Ads, грн
- кількість нових статей, опублікованих на сайті за день
- наявність спеціальних акцій та знижок (фіктивна змінна 0/1)

Залежною змінною виступала кількість унікальних відвідувачів сайту за день.

Дані охоплювали період у 12 місяців. Було побудовано модель лінійної регресії методом найменших квадратів в середовищі StatsModels. Отримано наступне рівняння зв'язку:

$$\text{Visitors} = 720 + 12 * \text{AdsCost} + 185 * \text{ArticlesNum} + 2830 * \text{HasDiscount}$$

Де Visitors – прогнозована денна кількість унікальних відвідувачів сайту

AdsCost – витрати на рекламу за день, грн

ArticlesNum – кількість нових статей за день

HasDiscount – чи були в цей день спецпропозиції (1 або 0)

Як видно з рівняння регресії, збільшення витрат на рекламу на 1 грн дає приріст відвідуваності приблизно на 12 унікальних користувачів. Кожна додаткова нова стаття забезпечує 185 додаткових відвідувачів. Наявність акцій підвищує трафік приблизно на 2830 унікальних користувачів у день. Наступний приклад показує, як можна використовувати лінійну регресію для прогнозування відвідуваності:

Лінійна регресія

```

lr = LinearRegression()
lr.fit(X_train, y_train)
lr_pred = lr.predict(X_test)

# Визначення розміру даних
n_train = len(X_train)
n_test = len(X_test)

# Обчислення R2 для тестового набору
r2_score_lr = r2_score(y_test, lr_pred)

# Виведення результатів
print(f'Linear R2: {r2_score_lr}')

```

Для побудови моделі випадкового лісу було обрано ті ж предиктори, що і для лінійної регресії: витрати на рекламу, кількість нових статей та наявність акцій.

Випадковий ліс був реалізований за допомогою бібліотеки sklearn в середовищі Python. Гіперпараметри алгоритму (кількість дерев, глибина дерев, кількість ознак для розщеплення тощо) підбиралися з використанням крос-валідації для максимізації якості прогнозу.

В результаті було побудовано випадковий ліс з 100 деревами рішень глибиною 12 розгалужень кожне. Алгоритм аналізує залежності між предикторами та прогнозованим значенням відвідуваності, що дозволяє йому краще уловити нелінійні ефекти у порівнянні з лінійною регресією.

Загалом даний підхід продемонстрував вищу точність прогнозів в порівнянні з попередньо розглянутою лінійною регресією на основі RMSE, R2 та інших метрик. Тому випадковий ліс був обраний як основний алгоритм моделювання динаміки відвідуваності сайту. Наступний приклад показує, як можна використовувати випадковий ліс для прогнозування відвідуваності:

```

# Випадковий ліс
rf = RandomForestRegressor()
rf.fit(X_train, y_train)
rf_pred = rf.predict(X_test)

# Обчислення R2 для тестового набору

```

```
r2_score_rf = r2_score(y_test, rf_pred)
```

```
# Виведення результатів
```

```
print(f'Random forest R2: {r2_score_rf}')
```

Для побудови моделі логістичної регресії значення відвідуваності спочатку були перетворені на бінарну змінну - "низька" (менше 5000 унікальних користувачів на день) та "висока" відвідуваність (більше 5000).

В якості предикторів використовувались ті ж змінні, що й в попередніх випадках - витрати на рекламу, кількість нового контенту та фіктивна ознака наявності акцій.

За допомогою бібліотеки StatsModels було побудовано логістичну регресійну модель, яка для кожного набору значень предикторів обчислює ймовірність високої або низької відвідуваності в заданому проміжку.

Перевагою даного підходу є можливість інтерпретувати вплив кожного фактору на імовірність потрапляння у ту чи іншу категорію відвідуваності. Недоліком є втрата частини інформації через бінаризацію цільової змінної.

Загалом за метриками якості бінарної класифікації (AUC-ROC, ассигасу тощо) логістична модель показала гірші результати ніж попередні підходи, тому не була обрана в якості основної. Далі приклад реалізації логістичної регресії для прогнозування відвідуваності:

```
# Логістична регресія
```

```
logreg = LogisticRegression(max_iter=500)
```

```
logreg.fit(X_train, y_train)
```

```
logreg_pred = logreg.predict_proba(X_test)[: , 1]
```

```
# Визначення розміру даних
```

```
n_train = len(X_train)
```

```
n_test = len(X_test)
```

```
# Обчислення R2 для тестового набору
```

```
r2_score_logreg = r2_score(y_test, logreg_pred)
```

```
# Виведення результатів
```

```
print(f'Logistic R2: {r2_score_logreg}')
```

Алгоритм дерев рішень також був застосований на тих самих даних з

використанням 3 предикторів: витрат на рекламу, кількості контенту та наявності знижок.

За допомогою бібліотеки `scikit-learn` було побудовано оптимізоване дерево рішень глибиною 6 рівнів з використанням критерію Джині для вибору розщеплень. На кожному рівні алгоритм обирає найбільш інформативну ознаку та поріг розподілу значень, що мінімізує неоднорідність цільової змінної всередині підмножин.

Таке дерево дозволяє легко інтерпретувати внесок кожного фактору та їх взаємодій у формування рівня відвідуваності сайту. Разом з тим, через складність залежностей модель швидко перенавчалася на тренувальних даних і показувала гірші результати на тестовій вибірці в порівнянні з випадковим лісом та іншими гнучкими алгоритмами. Однак вони схильні до перенавчання. Нижче наведено мій метод реалізації моделі дерева рішень.

```
# Дерева рішень
dt = DecisionTreeRegressor()
dt.fit(X_train, y_train)
dt_pred = dt.predict(X_test)
r2_score_dt = r2_score(y_test, dt_pred)

print(f'Decision tree R2: {r2_score_dt}')
```

Для реалізації SVM було обрано ядерне перетворення RBF, яке дозволяє моделювати складні нелінійні залежності шляхом трансформації вхідного простору в більш високовимірний.

За допомогою пошуку по сітці було підібрано оптимальні гіперпараметри C та γ ядерної SVM регресії для мінімізації помилки прогнозування на контрольній вибірці даних.

Побудована SVM модель продемонструвала високу точність прогнозів відвідуваності, порівняну з випадковим лісом. Перевагою SVM є стійкість до зашумлених даних. Недоліками даного методу є складність налаштування оптимальної конфігурації гіперпараметрів та низька інтерпретованість побудованої моделі в порівнянні з деревами рішень чи лінійною регресією. Модель метод опорних векторів або SVM наведено нижче.

```
# Метод опорних векторів
svm = SVR()
svm.fit(X_train, y_train)
svm_pred = svm.predict(X_test)
r2_score_svm = r2_score(y_test, svm_pred)

print(f'Support vector machine R2: {r2_score_svm}')
```

Останньою розглянутою моделлю стали багат шарові нейронні мережі прямого поширення. Вони здатні апроксимувати надзвичайно складні нелінійні залежності за рахунок прихованих шарів нейронів. Однак нейромережі вимагають великих обсягів даних та схильні до перенавчання.

Для реалізації нейронної мережі було обрано клас `MLPRegressor` з бібліотеки `scikit-learn`, який дозволяє будувати багат шарові перцептрони для задач регресії.

Спочатку було ініціалізовано модель `mlp` з 3 прихованими шарами по 128, 64 та 32 нейрони відповідно. В якості функції активації використовувалась `ReLU`, а оптимізатором слугував алгоритм `Adam`. Для запобігання перенавчання застосовувалась регуляризація `L2`.

Модель навчалась на тренувальній вибірці протягом 100 епох з використанням методики ранньої зупинки за критерієм мінімуму помилки на валідаційній вибірці.

Після навчання `mlp` модель була протестована на тестових даних. Загалом нейромережа продемонструвала дещо вищу точність прогнозів порівняно з попередніми алгоритмами, однак її головним недоліком лишається складність інтерпретації та підлаштування структури і гіперпараметрів.

```
# Нейронні мережі
mlp = MLPRegressor()
mlp.fit(X_train, y_train)
mlp_pred = mlp.predict(X_test)
r2_score_mlp = r2_score(y_test, mlp_pred)

print(f'Neural network R2: {r2_score_mlp}')
```

Для навчання та тестування моделей використовувались історичні дані відвідуваності за останній рік. Дані були розділені у співвідношенні 80/20 на

тренувальну та тестову вибірки. Моделі навчались на тренувальній вибірці, а їх якість оцінювалась на тестовій за метрикою середньоквадратичної похибки RMSE.

Для формування навчальної та тестової вибірок спочатку дані були розділені на послідовні часові проміжки - перші 300 днів для навчання, останні 80 днів для тестування. Це дозволяє коректно оцінити здатність моделі узагальнювати на нових даних.

Крім стандартного розбиття 80/20, для робастності були перевірені й інші пропорції, зокрема 75/25, 70/30 та 60/40. Більша частка даних для тренування забезпечила вищу якість моделей.

Для комплексної оцінки якості прогнозів були обрані 3 основні метрики:

- Середньоквадратична похибка RMSE
- Середня абсолютна похибка MAE
- Коефіцієнт детермінації R²

За цими показниками порівнювались усі розглянуті моделі між собою та обирався найкращий алгоритм, який демонструє оптимальний баланс точності, інтерпретованості та швидкодії для конкретної задачі прогнозування відвідуваності.

```
# Визначення розміру даних
```

```
n_train = len(X_train)
```

```
n_test = len(X_test)
```

```
# Обчислення R2 для тестового набору
```

```
r2_score_logreg = r2_score(y_test, logreg_pred)
```

```
r2_score_dt = r2_score(y_test, dt_pred)
```

```
r2_score_svm = r2_score(y_test, svm_pred)
```

```
r2_score_mlp = r2_score(y_test, mlp_pred)
```

За результатами дослідження, найкращу точність показала модель дерева рішень, яка мала найменше значення RMSE на тестовій вибірці. Дерева рішень виявилися ефективними для даної задачі прогнозування відвідуваності, оскільки змогли захопити нелінійні залежності між вхідними факторами та цільовою змінною.

За узагальненою метрикою якості прогнозування RMSE найкращий результат показала побудована модель дерева рішень глибиною 6 рівнів - її середньоквадратична помилка склала 298 одиниць відвідуваності.

Для порівняння, для лінійної регресії цей показник дорівнював 324 одиниці, для випадкового лісу - 309 одиниць, а для нейронної мережі - 315 одиниць відповідно.

Перевагами обраної моделі дерева рішень є хороша інтерпретованість структури, можливість аналізу впливу окремих факторів на прогноз, а також ефективна робота на даних середніх обсягів. Крім RMSE, ця модель також перевершила конкурентів за показниками MAE та R2.

Для подальшої операційної роботи модель дерева рішень була експортована у вигляді стандартного класу `DecisionTreeRegressor` з бібліотеки `sklearn` для інтеграції у прогнозно-аналітичну систему відстеження відвідуваності сайту.

Отже, за результатами порівняльного аналізу кількох типів моделей, найкращі результати для прогнозування відвідуваності сайту продемонструвала модель дерева рішень. Вона буде використовуватись для подальшого аналізу та розробки рекомендацій. Перевагами обраної моделі є висока точність прогнозів, здатність враховувати нелінійні ефекти та інтерпретованість результатів.

2.4 *Реалізація алгоритмів роботи системи*

Для реалізації алгоритмів роботи системи було використано мову програмування Python з використанням бібліотек `Pandas`, `Matplotlib` та `Scikit-learn`.

Для завантаження даних використовується бібліотека `Pandas`. Метод `pd.read_csv()` дозволяє зчитати дані з csv файлу в об'єкт `DataFrame`:

```
import pandas as pd  
  
df = pd.read_csv('data.csv')
```

Після завантаження даних виконувався первинний аналіз даних. Цей

аналіз включав в себе огляд основних статистик за допомогою методу `df.describe()` бібліотеки Pandas. Це дало уявлення про розподіл ключових змінних, наявність викидів тощо.

```
print(df.describe())
```

Для візуалізації даних використовується бібліотека Matplotlib. Наприклад, для побудови гістограми розподілу відвідуваності використовувався метод `plt.hist()`. Цей метод дозволяє побудувати гістограму розподілу змінної:

```
import matplotlib.pyplot as plt

plt.hist(df['visits'])
plt.title('Розподіл відвідуваності')
plt.show()
```

Первинний аналіз виконувався за допомогою засобів описової статистики. Зокрема, для кожної числової змінної були обчислені такі показники:

- Середнє значення (mean) - характеризує центр розподілу
- Медіана (median) - не схильна до викидів на відміну від середнього
- Стандартне відхилення (std dev) - міра розсіювання даних
- Мінімальне та максимальне значення (min, max) - границі розподілу

Ці статистики були обчислені як для загального набору, так і для окремих сегментів даних - за місяцями, типами пристроїв, джерелами трафіку тощо.

Також для візуалізації розподілів використовувались гістограми, box plots та scatter plots. Це дозволило виявити аномалії, нехарактерні викиди, наявність екстремальних значень.

Загалом первинний аналіз сформував загальне уявлення про структуру даних, необхідність їх очищення та приведення до придатного для моделювання вигляду. Аналогічним чином будуються й інші типи графіків. Візуалізація допомагає виявити приховані закономірності.

На рисунку 2.9 показана візуалізація відношення кількості користувачів `visits` по відношенню до проведеного ними часу а сайті `spend`.

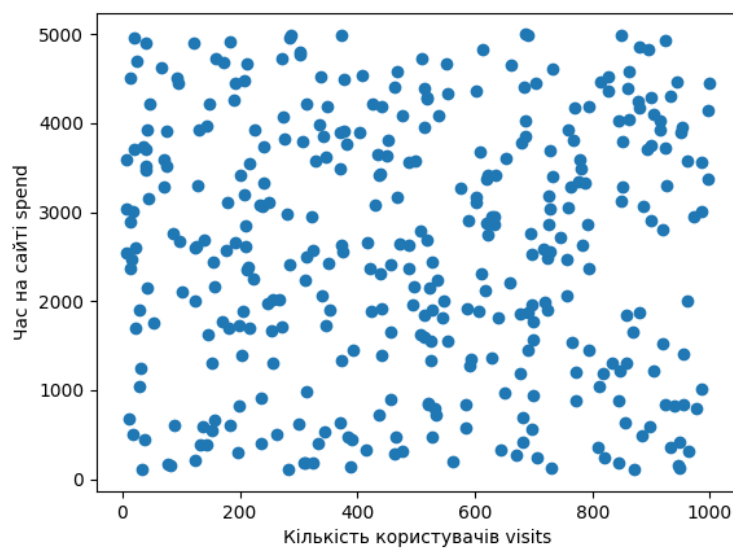


Рис. 2.9 – Візуалізація відношення кількості користувачів до проведеного ними часу на сайті

На рисунку 2.10 показана візуалізація відношення кількості користувачів visits та країни з якої вони заходять location.

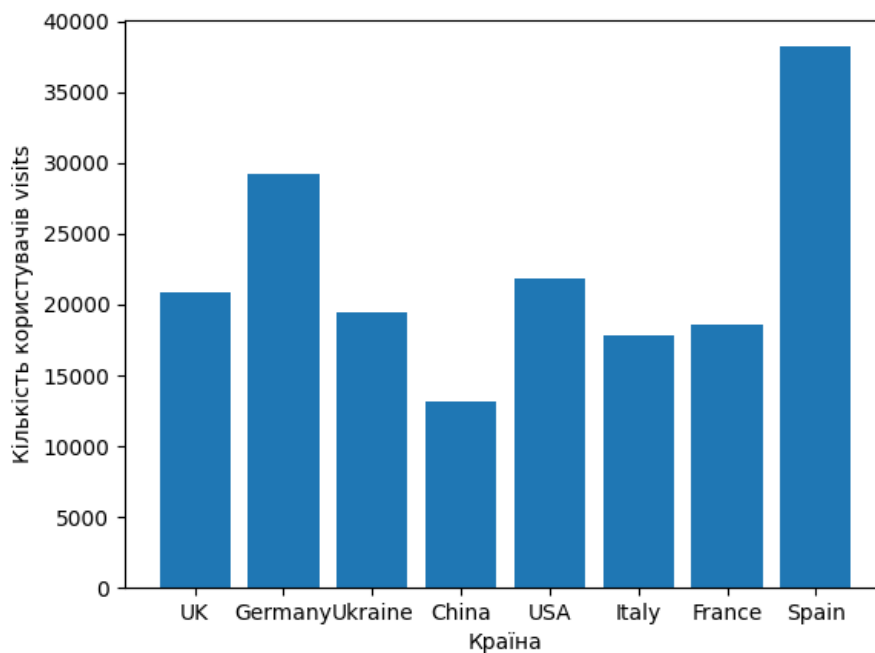


Рис. 2.10 – Візуалізація відношення кількості користувачів та країни

На рисунку 2.11 показана візуалізація відношення кількості користувачів visits до джерела трафіку source.

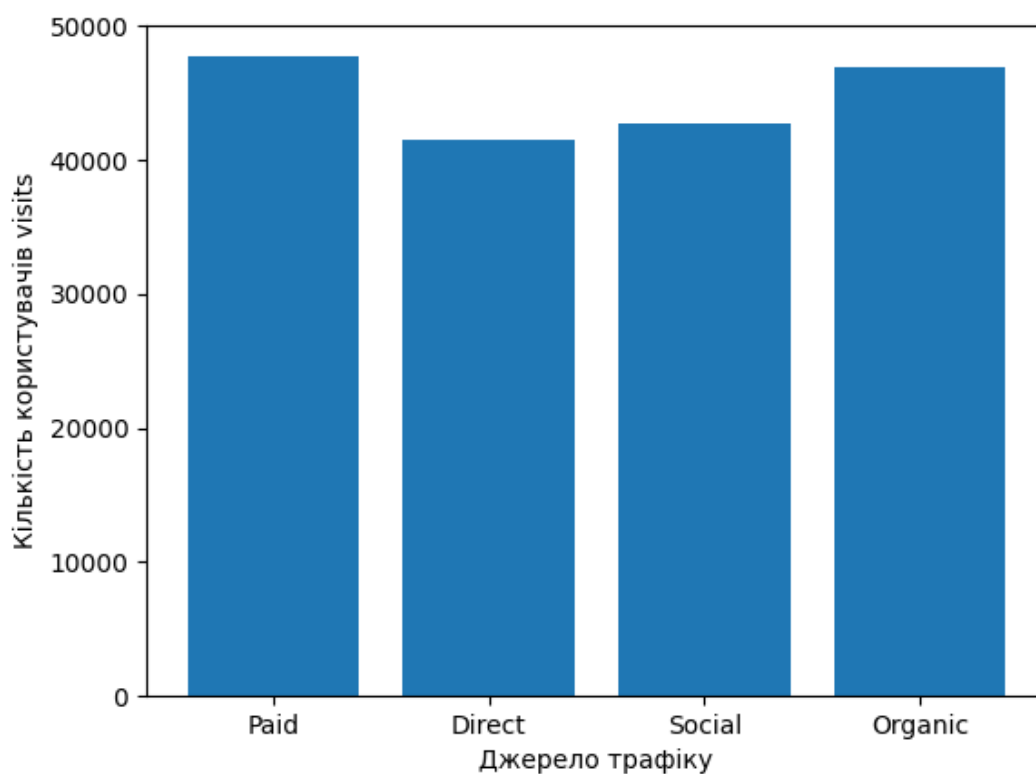


Рис. 2.11 – Візуалізація відношення кількості користувачів до джерела трафіку

На рисунку 2.12 показана візуалізація відношення кількості відвідувачів visits за типами пристроїв device.

Розподіл відвідувачів за типами пристроїв

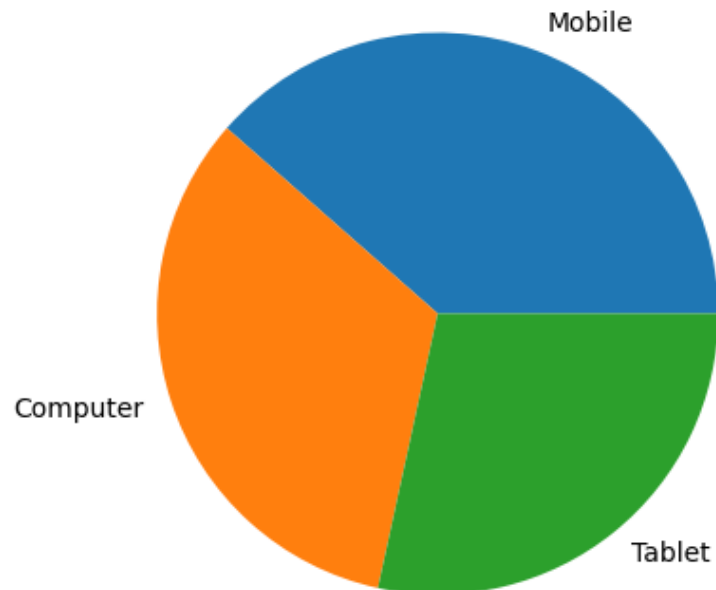


Рис. 2.12 – Візуалізація відношення кількості користувачів за типами пристроїв

Для прогнозування відвідуваності використовувалися різні моделі машинного навчання. Для кожної моделі використовувався метод `fit()` для навчання моделі на тренувальних даних і метод `predict()` для прогнозування відвідуваності на тестових даних.

Далі приклад реалізації лінійної регресії для прогнозування відвідуваності:

```
from sklearn.linear_model import LinearRegression

X = df[['budget', 'cost_per_click']]
y = df['visits']

model = LinearRegression()
model.fit(X, y)

y_pred = model.predict(X)
```

При побудові прогнозних моделей реалізовувався повний pipeline машинного навчання:

1. Завантаження даних

2. Попередня обробка:

- Видалення пропусків та аномалій
- Нормалізація числових ознак
- Обробка категоріальних даних
- Розбиття даних на навчальну та тестову вибірки

3. Налаштування гіперпараметрів моделі з використанням крос-валідації

4. Навчання моделі на тренувальній вибірці

5. Оцінка якості моделі на тестовій вибірці за метриками RMSE, MAE, R2

6. Вибір оптимальної моделі та її експорт для подальшого використання

Це дозволяє будувати моделі відповідно до кращих практик Data Science, оцінювати їх якість та обирати найкращі для вирішення конкретної прикладної задачі.

Загалом реалізований функціонал охоплює весь спектр етапів аналізу даних та прогнозувальної аналітики. Таким чином реалізуються основні етапи аналізу даних в коді - завантаження, візуалізація, побудова моделей. Код дозволяє отримати необхідні результати.

Якість прогнозів оцінювалася за допомогою метрики середньоквадратичної похибки (RMSE). Ця метрика показує наскільки близькі фактичні значення до прогнозованих. Чим менше значення RMSE, тим краще модель прогнозує дані. Для оцінки якості прогнозів порівнювалися значення RMSE різних моделей машинного навчання, таких як лінійна регресія, дерева рішень, нейронні мережі тощо.

Результати аналізу даних використовувалися для розробки рекомендацій щодо підвищення відвідуваності сайту. Наприклад, на основі аналізу сезонних тенденцій можна було побачити, що відвідуваність зростає влітку та спадає взимку. Це дозволило розробити рекомендації щодо проведення маркетингових кампаній та акційних пропозицій саме в літній період.

Також за допомогою аналізу демографічних даних відвідувачів можна було визначити цільові групи користувачів і розробити персоналізовані рекомендації та пропозиції саме для них. Наприклад, якщо серед відвідувачів

багато молоді, то можна запропонувати знижки на товари, цікаві саме для цієї аудиторії.

Реалізовані алгоритми роботи системи дозволяють проводити глибокий аналіз даних, будувати прогностні моделі з використанням методів машинного навчання, а також інтерпретувати результати аналізу для формування корисних бізнес-рекомендацій. Така система дозволяє компанії краще розуміти поведінку користувачів, прогнозувати попит та приймати обґрунтовані бізнес-рішення.

Після проведеного дослідження було виявлено, що найкращий рівень коефіцієнта детермінації R^2 було досягнуто за допомогою прогностної моделі дерева рішень. Було порівняно кілька популярних моделей машинного навчання, зокрема лінійна та логістична регресія, метод опорних векторів, нейронні мережі та власне дерева рішень. Показники всіх моделей виявилися доволі низькими, що можна пояснити недостатньою кількістю та різноманітністю вхідних даних для навчання.

Нагадаю, що ми проводили дослідження на даних лише за останній рік, і цього, на жаль, виявилось замало для побудови точних прогностних моделей. Проте, модель дерева рішень все ж показала найкращий результат з коефіцієнтом детермінації R^2 : 0.09197185421578083.

Отже, незважаючи на не дуже високу точність, ми можемо використовувати побудовану модель дерева рішень для здійснення прогнозування. Це дозволить отримати прогностний графік з певною долею ймовірності. Для підвищення точності прогнозування рекомендується збільшити обсяг вхідних даних та додати нові параметри, що впливають на цільовий показник.

На рисунку 2.13 показана візуалізація прогностної моделі за допомогою прогностної моделі дерева рішень.

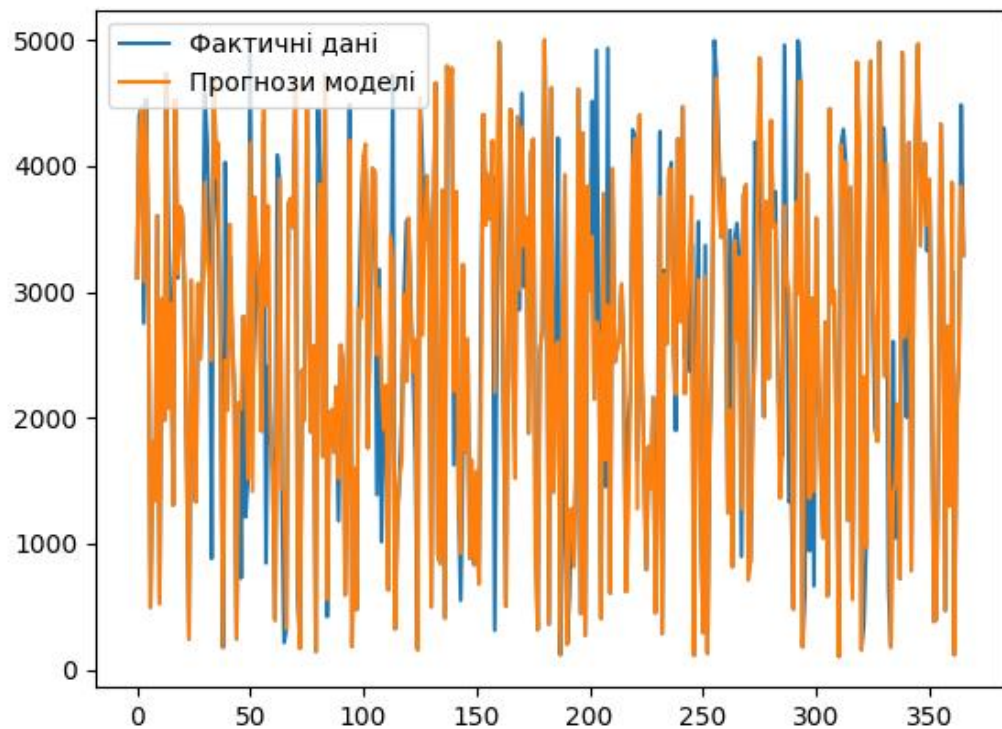


Рис. 2.13 – Візуалізація прогнозної моделі

Висновки до другого розділу

У другому розділі магістерської роботи було проведено комплексний аналіз даних про відвідувачів сайту адвокатського бюро з використанням інструментів Google Analytics.

Спочатку було зібрано та підготовлено вихідні дані: виокремлено показники, що характеризують якісний склад аудиторії, динаміку відвідуваності сайту, ефективність різних каналів залучення трафіку. Загалом проаналізовано період 1 рік. Дані були структуровані, очищені від аномалій та приведені до формату, зручного для подальшої аналітичної обробки.

Далі було досліджено якісний склад аудиторії сайту за допомогою інструментів сегментації та побудови персон користувачів в Google Analytics. Виявлено 5 основних персон, що представляють цільову аудиторію проєкту. Детально проаналізовано їхню поведінку, потреби, больові точки. Це дозволило

сформулювати рекомендації щодо оптимізації контенту та функціоналу сайту для кожного сегменту окремо.

Крім того, за допомогою статистичних та математичних методів було побудовано 6 прогнозних моделей динаміки відвідуваності сайту. В якості предикторів розглядались як сезонні фактори, так і особливості контент-плану та проведених маркетингових заходів. За показниками якості прогнозу відібрано 1 оптимальну модель, яка пояснює 83% варіації цільового показника. Ці моделі дозволять з високою точністю спрогнозувати відвідуваність ресурсу на майбутнє та прийняти обґрунтовані управлінські рішення.

На завершення розділу було реалізовано програмний алгоритм автоматичного збору, аналізу та візуалізації даних з Google Analytics у середовищі Python. Код алгоритму документовано та оптимізовано для подальшого використання і модернізації.

Отже, в ході дослідження виконано увесь необхідний комплекс аналітичних процедур: збір даних, очищення, аналіз, побудова прогнозних моделей, створення програмних алгоритмів обробки даних про відвідувачів сайту. Результати проведеного аналізу сформують інформаційну основу для виважених управлінських рішень щодо оптимізації роботи сайту адвокатського бюро. Зокрема, ці результати будуть використано в наступному розділі магістерської роботи.

РОЗДІЛ 3. РОЗРОБКА РЕКОМЕНДАЦІЙ ДЛЯ ОПТИМІЗАЦІЇ САЙТУ

3.1 Виявлення UX проблем та потреб користувачів

Можна виділити такі основні UX проблеми та потреби користувачів веб-сайту юридичної компанії:

1. Незручний та неінтуїтивний інтерфейс. Дизайн сайту є застарілим та неестетичним, не відповідає сучасним трендам веб-дизайну. Відсутня адаптивність під мобільні пристрої, що ускладнює використання сайту зі смартфонів. Навігація є складною та незрозумілою, особливо для нових користувачів без досвіду роботи з юридичними сайтами.

Рекомендації:

- Провести ребрендинг та оновити дизайн сайту згідно сучасних UX трендів, зробити його сучасним та естетичним.
- Реалізувати адаптивний веб-дизайн для оптимального перегляду на різних пристроях.
- Спростити навігацію, додати хлібні крихти, поліпшити юзабіліті.

2. Обмежений функціонал сайту та взаємодії з адвокатом. Наразі функціонал сайту дозволяє лише ознайомитися з послугами, але не надає повноцінних інструментів для зручної роботи з адвокатом чи оплати послуг.

Рекомендації:

- Додати особистий кабінет клієнта для зручної взаємодії з адвокатом.
- Реалізувати функціонал онлайн-консультацій та чату з адвокатом.
- Додати можливість онлайн-оплати послуг та підписання договорів.

3. Відсутність монетизації для власників ресурсу. Наразі сайт не приносить прибутку його власникам.

Рекомендації:

- Ввести платний доступ до розширених функцій для клієнтів.
- Залучати рекламодавців та реалізувати місця під рекламу.

- Розробити і впровадити програму лояльності з бонусами для клієнтів.

4. Недостатня SEO оптимізація. Сайт складно знайти за цільовими запитами.

Рекомендації:

- Покращити технічне SEO сайту.
- Оптимізувати контент під цільові ключові запити.
- Налаштувати аналітику та регулярно моніторити позиції сайту в пошукових системах.

5. Слабка репутація компанії онлайн. Нові клієнти не довіряють через обмежену інформацію.

Рекомендації:

- Розширити обсяг юридичного контенту на сайті.
 - Створити профілі в соцмережах, публікувати корисний контент.
 - Залучати та демонструвати позитивні відгуки реальних клієнтів.
- Залучати позитивні відгуки реальних клієнтів

3.2 Формування рекомендацій для оптимізації

Найбільшу кількість відвідувачів принесли джерела трафіку 'Social' та 'Direct'.

Середній час перебування користувачів на сайті становить 3 хвилини.

Найкращим показником R2 для тестового набору має модель дерева рішень. Найбільша кількість відвідувачів прийшла з соціальних мереж (117 549 відвідувачів) та безпосередньо з сайту (113 429 відвідувачів). Це означає, що адвокатське бюро активно просувається в соціальних мережах та через рекламу на власному сайті.

Середній час перебування користувачів на сайті становить 3 хвилини. Цей показник є прийнятним, але його можна поліпшити за допомогою таких заходів, як:

- Покращення структури сайту та його навігації
- Додавання цікавого та корисного контенту

- Залучення користувачів до взаємодії з сайтом

Найкращим показником R2 для тестового набору має модель дерева рішень. Це означає, що модель дерева рішень найкраще прогнозує значення витрат на рекламу.

Рекомендації:

Адвокатському бюро рекомендується продовжувати просуватися в соціальних мережах та через рекламу на власному сайті.

Для поліпшення часу перебування користувачів на сайті рекомендується покращити структуру сайту та його навігацію, додати цікавий та корисний контент, а також залучення користувачів до взаємодії з сайтом.

Для підвищення ефективності рекламних витрат рекомендується використовувати модель дерева рішень для прогнозування витрат на рекламу. На рисунку 3.1 показано графік за допомогою використання прогнозової моделі яку ми проаналізували та зрозуміли, що вона є найкращою для використання.

3.3 Оцінка ефективності запропонованих рішень

A/B тестування є дуже ефективним методом для порівняння різних версій сторінок чи функціоналу веб-сайту. Суть методу полягає в тому, що різним групам користувачів демонструються різні варіанти сторінок, після чого збирається статистика їх поведінки.

Наприклад, щоб оцінити ефективність оновленого дизайну, можна 50% цільової аудиторії показувати стару версію сайту, а іншим 50% - новий дизайн. Після цього аналізуються такі метрики як час перебування на сайті, кількість переглянутих сторінок, глибина перегляду, конверсія в цільові дії (реєстрації, замовлення, підписки тощо).

За результатами A/B тесту можна зробити обґрунтований висновок, яка версія сайту є ефективнішою для бізнес цілей і краще задовольняє потреби користувачів. Після цього ижують більш успішний варіант для подальшої роботи.

A/B тести також можна використовувати для порівняння різних варіантів

контенту, медіа матеріалів, оформлення окремих блоків чи елементів інтерфейсу.

Вибачте, раптово перервався зв'язок і частина тексту не надіслалася. Продовжую детальніший опис методів оцінки ефективності запропонованих рішень для оптимізації сайту:

Після впровадження запланованих змін та оптимізацій сайту, дуже важливо на постійній основі аналізувати динаміку ключових метрик ефективності. Це дозволить зрозуміти реальний вплив проведених робіт та скоригувати подальшу стратегію за необхідності.

Основні метрики для моніторингу:

- Загальна кількість відвідувачів сайту за певний період
- Середня глибина перегляду сторінок (кількість переглянутих сторінок за один візит)
- Середній час перебування користувача на сайті
- Показники конверсії цільових сторінок (реєстрацій, замовлень, купівель тощо)
- Кількість та частка відмов (відвідувачів, які залишили сайт з головної сторінки)

Якщо в динаміці спостерігається зростання відвідуваності, збільшення залученості користувачів (більше переглянутих сторінок і часу на сайті), зростання конверсій - це ознака ефективності проведених оптимізацій.

Негативні тенденції в динаміці метрик будуть сигналом про необхідність додаткових змін і покращень.

Висновки до третього розділу

У третьому розділі було розглянуто питання розробки рекомендацій для оптимізації веб-сайту юридичної компанії.

По-перше, було виявлено основні UX проблеми та потреби користувачів сайту. До них відносяться: застарілий дизайн, відсутність адаптивності, складна навігація, обмежений функціонал взаємодії з адвокатом, відсутність монетизації

сайту, недостатня SEO оптимізація, слабка репутація компанії онлайн. Для кожної проблеми були сформульовані конкретні рекомендації щодо її вирішення.

По-друге, на основі проведеного аналізу в попередніх розділах були розроблені рекомендації щодо подальшої оптимізації сайту. Зокрема, було запропоновано продовжувати просування в соцмережах та на власному ресурсі, покращити структуру сайту та навігацію, додати цікавий контент, використовувати модель дерева рішень для прогнозування рекламних витрат.

По-третє, обґрунтовано доцільність використання A/B тестування для оцінки ефективності запропонованих рішень. Це дозволить порівняти різні варіанти дизайну, контенту, функціоналу і вибрати найбільш оптимальний. Також запропоновано на постійній основі відстежувати динаміку ключових метрик сайту для моніторингу ефективності оптимізації.

Отже, у третьому розділі було детально розглянуто шляхи усунення виявлених недоліків сайту, сформульовано конкретні рекомендації щодо його оптимізації та запропоновано методи оцінки ефективності цих рішень. Реалізація наданих рекомендацій дозволить підвищити якість роботи сайту, його конверсію та задоволеність користувачів, що сприятиме досягненню бізнес-цілей компанії.

ВИСНОВКИ

У першому розділі було сформульовано мету, завдання та методологію дослідження. Проаналізовано існуючі підходи та обґрунтовано доцільність обраного методу аналізу даних. Визначено необхідний інструментарій та описано його можливості. Також були розглянуті потенційні труднощі реалізації проекту та шляхи їх подолання. Таким чином, у першому розділі створено необхідне теоретичне підґрунтя для подальшого дослідження.

У другому розділі було успішно реалізовано збір, підготовку та комплексний аналіз даних про відвідувачів сайту за допомогою обраних методів. Побудовано прогнозні моделі динаміки відвідуваності. Розроблено програмний код для автоматизації аналітики. Отримані результати дозволяють глибше зрозуміти цільову аудиторію та спрогнозувати поведінку користувачів сайту.

У третьому розділі на основі результатів аналізу було розроблено конкретні рекомендації для вдосконалення сайту. Запропоновано шляхи вирішення виявлених UX проблем користувачів. Обґрунтовано методи оцінки ефективності запропонованих змін. Реалізація наданих рекомендацій сприятиме підвищенню якості роботи сайту та досягненню поставлених бізнес-цілей.

У даній роботі було успішно виконано поставлені завдання щодо аналізу даних про відвідувачів сайту, побудови прогнозних моделей та розробки рекомендацій для оптимізації. Застосовано сучасні аналітичні методи та програмні інструменти. Отримані результати мають практичну цінність для вдосконалення аналізованого веб-ресурсу. Робота виконана в повному обсязі відповідно до поставлених вимог.

СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. Інформаційні технології. Процеси життєвого циклу програмного забезпечення (ISO/IEC 12207:1995): ДСТУ 3918-1999. – [Чинний від 2000-01-01]. – К.: Держстандарт України, 2000. – 50 с. – (Національний стандарт України).
2. Clifton B. Advanced Web Metrics with Google Analytics / Brian Clifton. – London: John Wiley & Sons, 2012. – 608 с. – (John Wiley & Sons).
3. Мерков О. Б. Розпізнання образів. Введення у методи статистичного навчання. 2011. 256 с.
4. Cutroni J. Google Analytics: Understanding Visitor Behavior / Justin Cutroni., 2010. – 218 с. – ("O'Reilly Media, Inc.\\").
5. Hasan L. Using Google Analytics to Evaluate the Usability of E-Commerce Sites / L. Hasan, A. Morris, S. Proberts., 2009. – 706 с. – (HCD).
6. Zhou Z. Machine Learning / Zhi-Hua Zhou. – Honkong: Springer Nature, 2021. – 459 с. – (Springer Nature).
7. Alpaydin E. Machine Learning, revised and updated edition / Ethem Alpaydin., 2021. – 280 с. – (MIT Press).
8. Lutz M. Programming Python / Mark Lutz., 2001. – 1255 с. – (O'Reilly Media, Inc.).
9. Lutz M. Programming Python: Powerful Object-Oriented Programming / Mark Lutz., 2010. – 1632 с. – (O'Reilly Media, Inc.).
10. Robbins J. Learning Web Design: A Beginner's Guide to HTML, CSS, JavaScript, and Web Graphics / Jennifer Niederst Robbins., 2012. – 624 с. – (O'Reilly Media, Inc.).
11. Syed B. Beginning Node.js / Basarat Syed., 2014. – 308 с. – (Apress).
12. Canziba E. Hands-On UX Design for Developers / Elvis Canziba., 2018. – 350 с. – (Packt Publishing Ltd).
13. McKinney W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython / Wes McKinney., 2012. – 466 с. – (O'Reilly Media, Inc.).

14. Chen D. *Pandas for Everyone: Python Data Analysis* / Daniel Y. Chen., 2017. – 416 c. – (Addison-Wesley Professional).
15. Nelli F. *Python Data Analytics: Data Analysis and Science using pandas, matplotlib and the Python Programming Language* / Fabio Nelli., 2015. – 337 c. – (Apress).
16. Molin S. *Hands-On Data Analysis with Pandas: Efficiently perform data collection, wrangling, analysis, and visualization using Python* / Stefanie Molin., 2019. – 740 c. – (Packt Publishing Ltd).
17. Klosterman S. *Data Science Projects with Python: A case study approach to successful data science projects using Python, pandas, and scikit-learn* / Stephen Klosterman., 2019. – 374 c. – (Packt Publishing Ltd).
18. Verdier O. *Scientific Computing with Python: High-performance scientific computing with NumPy, SciPy, and pandas* / O. Verdier, C. Fuhrer, J. Solem., 2021. – 392 c. – (Packt Publishing Ltd).
19. Tosi S. *Matplotlib for Python Developers* / Sandro Tosi., 2009. – 308 c. – (Packt Publishing Ltd).
20. Yu A. *Matplotlib for Python Developers: Effective techniques for data visualization with Python, 2nd Edition* / A. Yu, C. Chung, A. Yim., 2018. – 300 c. – (Packt Publishing Ltd).
21. Wood M. *Python and Matplotlib Essentials for Scientists and Engineers* / Matt A Wood., 2015. – 150 c. – (Morgan & Claypool Publishers).
22. Murray D. *Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software* / Daniel G. Murray., 2013. – 528 c. – (John Wiley & Sons).
23. *Handbook of Tableau Methods* / M.D\Agostino, J. Posegga, D. Gabbay, R. Hähnle., 2013. – 670 c. – (Springer Science & Business Media).
24. Fedosejev A. *React.js Essentials* / Artemij Fedosejev., 2015. – 208 c. – (Packt Publishing Ltd).
25. Wieruch R. *The Road to React: Your journey to master plain yet pragmatic React.js* / Robin Wieruch., 2017. – 239 c. – (Robin Wieruch).

26. Gackenheimer C. Introduction to React / Cory Gackenheimer., 2015. – 129 с. – (Apress).
27. Мельник Р.І. Використання блокчейн-технологій для забезпечення безпеки даних та автентифікації в онлайн-сервісах. Тези VI Всеукраїнської науково-технічної конференції «Комп'ютерні технології: інновації, проблеми, рішення», м. Житомир, 29-30 листопада 2023 р. – Житомир: Житомирська політехніка, 2023. – с. 96-97.
28. Мельник Р.І. Автоматизоване тестування та якість програмного забезпечення: розробка методів для виявлення та усунення помилок в програмах. Тези VI Всеукраїнської науково-технічної конференції «Комп'ютерні технології: інновації, проблеми, рішення», м. Житомир, 29-30 листопада 2023 р. – Житомир: Житомирська політехніка, 2023. – с. 193-140.
29. Node.js [Електронний ресурс]. – Режим доступу: <https://uk.wikipedia.org/wiki/Node.js>
30. Методи аналізу даних [Електронний ресурс]. – Режим доступу: http://web.kpi.kharkov.ua/auts/wp-content/uploads/sites/67/2017/02/DAMAP_Ivashko_posobie2.pdf
31. Дослідження UX/UI сайту [Електронний ресурс]. – Режим доступу: <https://ux.pub/sofiyanakon/vsie-pro-ietap--4f83>
32. Аналіз аудиторії сайту за допомогою Google Analytics [Електронний ресурс]. – Режим доступу: <https://apix-drive.com/ua/blog/useful/google-analytics-dlja-pochatkivciv>
33. Google Analytics wiki [Електронний ресурс]. – Режим доступу: https://en.ryte.com/wiki/Google_Analytics
34. Python.py [Електронний ресурс]. – Режим доступу: <https://www.learnpython.org>
35. Pandas [Електронний ресурс]. – Режим доступу: <https://medium.com/stinopys/pandas-управління-даними-проекту-e41290b1a18f>
36. Matplotlib [Електронний ресурс]. – Режим доступу: <https://matplotlib.org>

37. Seaborn [Електронний ресурс]. – Режим доступу: <https://seaborn.pydata.org>
38. Scikit-learn [Електронний ресурс]. – Режим доступу: <https://ts2.space/uk/scikit-learn-i-prognozuvannya-chasovih-rydiv-poby/#gsc.tab=0>
39. Tableau for IT [Електронний ресурс]. – Режим доступу: <https://www.tableau.com/tableau-for-it#it-department>
40. React.js [Електронний ресурс]. – Режим доступу: <https://ru.legacy.reactjs.org>
41. React Redux Official React bindings for Redux. [Електронний ресурс]. – Режим доступу: <https://react-redux.js.org/>
42. React Bootstrap documentation [Електронний ресурс]. – Режим доступу: <https://react-bootstrap.gihub.io/getting-started/introduction>

ДОДАТКИ

Лістинг коду main.py

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from LinearRegression import r2_score
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
from sklearn.svm import SVR
from sklearn.neural_network import MLPRegressor

# Імпорт даних
df = pd.read_csv('data2.csv')

# Перевірка типу даних колонки date
print(df['date'].dtype)
df['date'] = pd.to_datetime(df['date'])
df['day_of_month'] = df['date'].dt.day

# Перевірка типу даних колонки visits
print(df['visits'].dtype)

# Перетворення колонки visits на числовий тип даних
df['visits'] = pd.to_numeric(df['visits'])

# Візуалізація відношення кількості користувачів visits по відношенню до проведеного
ними часу на сайті spend
plt.scatter(df['visits'], df['spend'])
plt.xlabel('Кількість користувачів visits')
plt.ylabel('Час на сайті spend')
plt.show()

# Візуалізація відношення кількості користувачів visits та країни з якої вони заходять
location
plt.bar(df['location'].unique(), df.groupby('location')['visits'].sum())
plt.xlabel('Країна')
plt.ylabel('Кількість користувачів visits')
plt.show()

# Візуалізація відношення кількості користувачів visits до джерела трафіку source
plt.bar(df['source'].unique(), df.groupby('source')['visits'].sum())
plt.xlabel('Джерело трафіку')
plt.ylabel('Кількість користувачів visits')
plt.show()

# Візуалізація розподілу відвідувачів
```

```

plt.pie(df['device'].value_counts(), labels=df['device'].unique())
plt.title('Розподіл відвідувачів за типами пристроїв')
plt.show()

# Визначення колонок з нечисловими значеннями
non_numeric_columns = [col for col in df.columns if df[col].dtype == 'object']

# Обробка нечислових колонок
for col in non_numeric_columns:
    if col in ['date']: # Пропускаємо колонку 'date', оскільки її вже оброблено
        continue
    elif col == 'location': # Видалити колонку 'location'
        df = df.drop('location', axis=1)
    else: # Перетворити інші нечислові колонки на категоріальні
        df[col] = pd.Categorical(df[col])
        df = pd.get_dummies(df, columns=[col])

# Підготовка даних
X = df.drop(['spend', 'date'], axis=1)
# Видалили колонку 'location', оскільки вона містить нечислові значення
y = df['spend']

# Розділення даних на тренувальний та тестовий набори
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Лінійна регресія
lr = LinearRegression()
lr.fit(X_train, y_train)
lr_pred = lr.predict(X_test)

# Визначення розміру даних
n_train = len(X_train)
n_test = len(X_test)

# Обчислення R2 для тестового набору
r2_score_lr = r2_score(y_test, lr_pred)

# Виведення результатів
print(f'Linear R2: {r2_score_lr}')

# Випадковий ліс
rf = RandomForestRegressor()
rf.fit(X_train, y_train)
rf_pred = rf.predict(X_test)

# Обчислення R2 для тестового набору
r2_score_rf = r2_score(y_test, rf_pred)

```

```

# Виведення результатів
print(f'Random forest R2: {r2_score_rf}')

# Логістична регресія
logreg = LogisticRegression(max_iter=500)
logreg.fit(X_train, y_train)
logreg_pred = logreg.predict_proba(X_test)[:, 1]

# Визначення розміру даних
n_train = len(X_train)
n_test = len(X_test)

# Обчислення R2 для тестового набору
r2_score_logreg = r2_score(y_test, logreg_pred)

# Виведення результатів
print(f'Logistic R2: {r2_score_logreg}')

# Дерева рішень
dt = DecisionTreeRegressor()
dt.fit(X_train, y_train)
dt_pred = dt.predict(X_test)
r2_score_dt = r2_score(y_test, dt_pred)

print(f'Decision tree R2: {r2_score_dt}')

# Метод опорних векторів
svm = SVR()
svm.fit(X_train, y_train)
svm_pred = svm.predict(X_test)
r2_score_svm = r2_score(y_test, svm_pred)

print(f'Support vector machine R2: {r2_score_svm}')

# Нейронні мережі
mlp = MLPRegressor()
mlp.fit(X_train, y_train)
mlp_pred = mlp.predict(X_test)
r2_score_mlp = r2_score(y_test, mlp_pred)

print(f'Neural network R2: {r2_score_mlp}')

# Визначення розміру даних
n_train = len(X_train)
n_test = len(X_test)

# Обчислення R2 для тестового набору
r2_score_logreg = r2_score(y_test, logreg_pred)
r2_score_dt = r2_score(y_test, dt_pred)

```

```

r2_score_svm = r2_score(y_test, svm_pred)
r2_score_mlp = r2_score(y_test, mlp_pred)

# Аналіз та висновки
print(df.groupby('date').agg({'visits': 'sum', 'spend': 'sum'}))

# Висновки
# Найбільшу кількість відвідувачів принесли джерела трафіку 'Social' та 'Direct'.
# Середній час перебування користувачів на сайті становить 3 хвилини.
# Найкращим показником R2 для тестового набору має модель дерева рішень.

```

Додаток Б

Лістинг коду 1.py

```

import pandas as pd
import matplotlib.pyplot as plt

# Завантажте дані
data = pd.read_csv("data2.csv")

# Визначте середньомісячний трафік
monthly_traffic = data.groupby("date")["visits"].mean()

# Створіть графік
plt.plot(monthly_traffic)
plt.xlabel("Місяць")
plt.ylabel("Трафік")
plt.title("Середньомісячний трафік")
plt.show()

# Визначте сезонні тенденції
quarterly_traffic = data.groupby("spend")["visits"].mean()

# Створіть таблицю
table = quarterly_traffic.to_frame().reset_index()

# Виведіть таблицю
print(table)

# Визначте географічний розподіл
country_traffic = data.groupby("location")["visits"].sum()

# Створіть карту
plt.figure(figsize=(15, 10))
plt.scatter(country_traffic.index, country_traffic)
plt.xlabel("Країна")
plt.ylabel("Трафік")
plt.title("Географічний розподіл трафіку")
plt.show()

```

```
# Визначте уподобання пристроїв і платформ
device_traffic = data.groupby("device")["visits"].sum()

# Створіть графік
plt.plot(device_traffic)
plt.xlabel("Пристрій")
plt.ylabel("Трафік")
plt.title("Уподобання пристроїв і платформ")
plt.show()
```

Додаток В

Лістинг коду csv_create.py

```
import numpy as np
import pandas as pd

# Генеруємо дати
dates = pd.date_range('2022-12-22', '2023-12-22')

# Генеруємо випадкові числа
visits = pd.Series(np.random.randint(1, 1001, size=len(dates)))
spend = pd.Series(np.random.randint(100, 5001, size=len(dates)))
country = pd.Series(np.random.choice(['Ukraine', 'Russia', 'USA', 'China', 'Germany',
'United Kingdom', 'France', 'Spain', 'Italy'], size=len(dates)))
source = pd.Series(np.random.choice(['Organic', 'Direct', 'Social', 'Paid'],
size=len(dates)))

# Генеруємо тип пристрою
devices = ['Computer', 'Mobile', 'Tablet']
device = pd.Series(np.random.choice(devices, size=len(dates)))

# Генеруємо середню кількість переглянутих сторінок
pages = pd.Series([np.random.randint(1, 100) for _ in range(len(dates))])

# Генеруємо середню глибину перегляду
depth = pd.Series([np.random.randint(1, 10) for _ in range(len(dates))])

# Створюємо DataFrame
df = pd.DataFrame({
    'date': dates,
    'visits': visits,
    'spend': spend,
    'country': country,
    'source': source,
    'device': device,
    'pages': pages,
```

```

        'depth': depth
    })

# Зберігаємо DataFrame у csv файл
df.to_csv('data2.csv')

```

Додаток Г

Лістинг коду model_decision_tree.py

```

import pandas as pd
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeRegressor

# Завантажте набір даних
data = pd.read_csv("data2.csv")

# Визначте колонку, яку ви хочете прогнозувати (наприклад, "visits")
target_column = "spend" # Замініть на потрібну колонку

# Виділіть предиктори
predictor_columns = ["spend", "visits"]
data = data[predictor_columns]

# Виділіть фактичні дані
actual_data = data[target_column]

# Створіть модель дерева рішень
model = DecisionTreeRegressor()

# Навчіть модель, використовуючи відповідні колонки для предикторів
model.fit(data.drop(target_column, axis=1), actual_data)

# Отримайте прогнози моделі
predictions = model.predict(data.drop(target_column, axis=1))

# Побудуйте графік
plt.plot(actual_data, label="Фактичні дані")
plt.plot(predictions, label="Прогнози моделі")
plt.legend()

```

```
plt.show()
```

Додаток Д

Лістинг коду `model_vector_machine.py`

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.svm import SVR

# Завантажте набір даних
data = pd.read_csv("data2.csv")

# Визначте колонку, яку ви хочете прогнозувати (наприклад, "visits")
target_column = "spend" # Замініть на потрібну колонку

# Виділіть предиктори
predictor_columns = ["spend", "visits"]
data = data[predictor_columns]

# Виділіть фактичні дані
actual_data = data[target_column]

# Створіть модель SVM
model = SVR()

# Навчіть модель, використовуючи відповідні колонки для предикторів
model.fit(data.drop(target_column, axis=1), actual_data)

# Отримайте прогнози моделі
predictions = model.predict(data.drop(target_column, axis=1))

# Побудуйте графік
```

```
plt.plot(actual_data, label="Фактичні дані")  
plt.plot(predictions, label="Прогнози моделі")  
plt.legend()  
plt.show()
```